# ON TESTING CONDITIONAL MOMENT RESTRICTIONS: THE CANONICAL CASE

GAUTAM TRIPATHI AND YUICHI KITAMURA

Abstract. Let $(x, z)$ be a pair of random vectors. We construct a new "smoothed" empirical likelihood based test for the hypothesis that $\mathbb{E}(z|x) \overset{a.s.}{=} 0$, and show that the test statistic is asymptotically normal under the null. An expression for the asymptotic power of this test under a sequence of local alternatives is also obtained. The test is shown to possess an optimality property in large samples. Simulation evidence suggests that it also behaves well in small samples.

## 1. Introduction

In a series of papers Owen (1988, 1990, 1991) studied the use of inference based on the nonparametric likelihood ratio. This approach is particularly useful when testing hypotheses that can be expressed as moment restrictions. As a specific example suppose that $\{z_1, \ldots, z_n\}$ is a random sample in $\mathbb{R}^d$, and we want to test the null hypothesis $\mathbb{E}z_1 = 0$. Owen's empirical likelihood ratio testing procedure is as follows: First, maximize the log likelihood under the null hypothesis of a discrete distribution that has support on the data; i.e. obtain the restricted empirical log likelihood

$$\mathrm{EL}^r = \max_{p_1,\ldots,p_n} \sum_{i=1}^n \log p_i \quad s.t. \quad p_i \geq 0,\ \sum_{i=1}^n p_i = 1,\ \sum_{i=1}^n z_i p_i = 0.$$

Next, obtain the unrestricted empirical log likelihood

$$\mathrm{EL}^{ur} = \max_{p_1,\ldots,p_n} \sum_{i=1}^n \log p_i \quad s.t. \quad p_i \geq 0,\ \sum_{i=1}^n p_i = 1.$$

Finally, construct the empirical likelihood ratio

$$\mathrm{ELR} = 2\{\mathrm{EL}^{ur} - \mathrm{EL}^r\}$$

and reject $\mathrm{H}_0$ if ELR is large. Owen demonstrated that critical values for this test can be obtained by using the fact that, under the null hypothesis, $\mathrm{ELR} \xrightarrow{d} \chi^2_d$ as $n \uparrow \infty$. As some recent papers (described later) have shown,

*Date*: This Version: September 20, 2000.
*Key words*: Conditional Moment Restrictions, Empirical Likelihood, Smoothing.
*Address correspondence to*: Gautam Tripathi, Department of Economics, University of Wisconsin, Madison, WI-53706. *E-Mail*: `gtripath@ssc.wisc.edu`.

this approach can be used to handle quite general forms of moment restrictions. However, the attention of most of the literature seems to have been confined to dealing with hypotheses expressed as *unconditional* moment restrictions. In this paper we extend the empirical likelihood paradigm to handle *conditional* moment restrictions.

Let $(x, z)$ denote a pair of observable random vectors. Throughout the paper we will treat $z$ as the response and $x$ as the conditioning variable. We extend the empirical likelihood approach to test

$$\mathrm{H}_0 : \Pr\{\mathbb{E}(z|x) = 0\} = 1 \quad vs. \quad \mathrm{H}_1 : \Pr\{\mathbb{E}(z|x) = 0\} < 1.$$

Note that this null hypothesis[1] is a special case of restrictions of the form $\mathbb{E}\{g(z, \theta_0)|x\} = 0$, where $\theta_0$ is an unknown finite dimensional parameter and $g$ a vector of known functions. However, the mathematical detail required in dealing with such restrictions is substantially higher than that involved in testing the prototypical conditional moment restriction $\mathbb{E}(z|x) = 0$. Therefore, our research program is to first obtain results for testing the canonical restriction $\mathbb{E}(z|x) = 0$. The results in this paper, apart from being theoretically interesting in their own right, are directly applicable to situations where we want to test for orthogonality of observed variables or to test the hypothesis of no relationship between the response and explanatory variable.

Much progress has been made in the area of testing conditional moment restrictions. See, among others, contributions by Newey (1985) and Bierens (1990, 1994). Related to this literature is the work on specification testing of a parametric regression function against a nonparametric alternative. See, for instance, the papers by Eubank and Spiegelman (1990), Wooldridge (1992), Yatchew (1992), Härdle and Mammen (1993), Whang and Andrews (1993), Fan and Li (1996), Hong and White (1995), Zheng (1996), Aït-Sahalia, Bickel, and Stoker (2000), Andrews (1997), Bierens and Ploberger (1997), and Horowitz and Spokoiny (1999). We show that a test based on Owen's empirical likelihood provides a useful alternative to the procedures developed in the above mentioned papers. Moreover, our test possesses an asymptotic optimality property and also appears to work well in finite samples.

The paper is organized is as follows: In Section 2 we introduce the smoothed empirical log likelihood (hereafter abbreviated as SEL) approach for testing $\mathbb{E}(z|x) = 0$. Section 3 describes the assumptions employed in this paper along with some notation which is used subsequently. Sections 4 and 5 have results regarding the asymptotic distribution of the SEL based test under the null, and under a sequence of local alternatives, respectively. In Section 6 we describe an optimality property of our test and in Section 7 we present some simulation results about the small sample behavior of our test. Section 8 concludes. All proofs are confined to the appendices.

[1]When writing conditional moment restrictions such as $\mathbb{E}(z|x) = 0$, we frequently omit the qualifier "w.p 1."

**Notation.** The following notation is used throughout the paper: By a "vector" we mean a column vector. We do not make any notational distinction between a random variable and the value taken by it. The difference should be clear from the context. $S$ is a subset of $\mathbb{R}^s$ which may be unbounded. When $S$ is open we let $\mathrm{C}^k(S)$ denote the set of all real valued functions on $S$ which have continuous partial derivatives up to order $k$. When $S$ is closed we say that $f \in \mathrm{C}^k(S)$ if $f \in \mathrm{C}^k(int(S))$ and $f$ along with its partial derivatives up to order $k$ can be extended continuously to $S$. $\mathrm{L}^2(S)$ stands for the Hilbert space of all real valued square integrable functions on $S$ which are integrable with respect to the probability distribution on $S$. $\mathbb{I}\{A\}$ is the indicator function of set $A$, and for a matrix $V$ the symbol $\|V\| = \sqrt{tr(VV')}$ denotes the Frobenius norm. $\|V\|$ reduces to the usual Euclidean norm in case $V$ happens to be a vector. Unless stated otherwise, all limits are taken as the number of observations $n \uparrow \infty$. To simplify typography we frequently suppress the dependence of a function upon $n$ without warning the reader. $\square$

## 2. The Smoothed Empirical Likelihood Approach

This section develops an empirical likelihood based test of conditional moment restriction $\mathbb{E}(z|x) = 0$. Our main tool is empirical likelihood, though a kernel smoothing technique plays an important part in formulating our test procedure. Recall that smoothing arises naturally in the theory of local likelihood estimation by considering expected log likelihood. See, for example, Brillinger (1977), Owen (1984), Hastie and Tibshirani (1986), and Staniswalis (1987). Our empirical likelihood ratio based test can also be motivated using an expected log likelihood criterion.

Let $f(x, z)$ denote the density of $(x, z)$ with respect to some appropriate measure. Define $f(x, z) = f(z|x)h(x)$, where $f(z|x)$ and $h(x)$ denote the conditional density of $z$ given $x$ and the marginal density of $x$, respectively. We want to test the conditional moment restriction $\int_{\mathbb{R}^d} z f(z|x)\, dF(z) = 0$, where $F$ is a dominating measure for the marginal distribution of $z$. Throughout the paper we assume that $x$ is continuously distributed and that $f(z|x)$ is smooth in $x$. To illustrate why smoothing the empirical likelihood is important, we first ignore the smoothness of $f(z|x)$ and see what happens when we calculate the empirical likelihood for this problem without any smoothing. So let $\{x_i, z_i\}_{i=1}^n$ be a random sample and $\nu_{x,n}$ and $\nu_{z,n}$ denote the counting measures on $\{x_i\}_{i=1}^n$ and $\{z_i\}_{i=1}^n$ respectively. Consider the $n+1$ sets of probability measures $\mathcal{P}^{(n)}_{z|x=x_i} = \{P_{z|x=x_i} \ll \nu_{z,n} : \int dP_{z|x=x_i} = 1, \int z\, dP_{z|x=x_i} = 0\}$ for $i = 1, \dots, n$, and $\mathcal{P}^{(n)}_x = \{P_x \ll \nu_{x,n} : \int dP_x = 1\}$. Let $p_{z_j|x_i}$ be the Radon-Nikodym derivative of $P_{z|x=x_i} \in \mathcal{P}^{(n)}_{z|x=x_i}$ with respect to $\nu_{z,n}$, evaluated at $(z_j, x_i)$. Here $i, j = 1, \dots, n$. Similarly, $p_{x_i}$ denotes the Radon-Nikodym derivatives of $P_x \in \mathcal{P}^{(n)}_x$ with respect to $\nu_{x,n}$, evaluated at $x_i$. Define $p_{x_i,z_i} = p_{z_i|x_i} p_{x_i}$. The conventional empirical likelihood is simply the multinomial likelihood $\Pi_{i=1}^n p_{x_i,z_i} = \Pi_{i=1}^n p_{z_i|x_i} p_{x_i}$ maximized over the Radon-Nikodym derivatives of $P_{z|x=x_i} \in \mathcal{P}^{(n)}_{z|x=x_i}$ and $P_x \in \mathcal{P}^{(n)}_x$. This is equivalent to maximizing

$$\sum_{i=1}^n \log p_{x_i,z_i} = \sum_{i=1}^n \log p_{z_i|x_i} + \sum_{i=1}^n \log p_{x_i} \tag{2.1}$$

with respect to $\{p_{z_j|x_i}, p_{x_i} : i, j = 1, \ldots, n\}$, subject to the constraints

$$p_{z_j|x_i} \geq 0,\ p_{x_i} \geq 0,\ \sum_{j=1}^n p_{z_j|x_i} = 1,\ \sum_{i=1}^n p_{x_i} = 1,\ \sum_{j=1}^n z_j p_{z_j|x_i} = 0. \tag{2.2}$$

Although maximizing (2.1) under (2.2) yields the usual nonparametric mle of $p_x$, namely $\hat{p}_{x_i} = 1/n$ for each $i$, it does not return a consistent solution for $p_{z_j|x_i}$. To see this, suppose $d = 1$ for simplicity and assume that the convex hull of $\{z_1, \ldots, z_n\}$ contains the origin so that the maximization problem is well defined. Then it is easy to see that the solution is to set at least $n-2$ of the $p_{z_j|x_i}$'s to zero, irrespective of the sample size. Moreover, without the last constraint in (2.2), the solution is $p_{z_j|x_i} = \delta_{ij}$, where $\delta_{ij}$ is Kronecker's delta. This, unfortunately, does not yield any meaningful results.

The above problem is analogous to the failure of likelihood based function estimation reported in Hastie and Tibshirani (1986, Section 5). The remedy they suggest is to maximize the expected log likelihood instead. Applying this idea to our problem, consider maximizing the empirical analog of

$$\mathbb{E}\{\log f(x,z)\} = \mathbb{E}\{\mathbb{E}[\log f(z|x)|x]\} + \mathbb{E}\{\log h(x)\}, \tag{2.3}$$

subject to (2.2). This leads to the following maximization problem:

$$\max_{\{p_{z_j|x_i}, p_{x_i} : i,j=1,\ldots,n\}} n^{-1} \sum_{i=1}^n \sum_{j=1}^n w_{ij} \log p_{z_j|x_i} + n^{-1} \sum_{i=1}^n \log p_{x_i} \tag{2.4}$$

$$s.t.\quad p_{z_j|x_i} \geq 0,\ p_{x_i} \geq 0,\ \sum_{j=1}^n p_{z_j|x_i} = 1,\ \sum_{i=1}^n p_{x_i} = 1,\ \sum_{j=1}^n z_j p_{z_j|x_i} = 0.$$

Here

$$w_{ij} = \frac{\mathcal{K}(\frac{x_i - x_j}{b_n})}{\sum_{j=1}^n \mathcal{K}(\frac{x_i - x_j}{b_n})} = \frac{\mathcal{K}_{ij}}{\sum_{j=1}^n \mathcal{K}_{ij}},$$

where the function $\mathcal{K}$ is chosen to satisfy Assumption 3.4. The $w_{ij}$'s are kernel weights familiar from the nonparametric regression literature and are mathematically quite tractable. The bandwidth $b_n$ is a null sequence of positive numbers satisfying certain conditions described later in the paper.

To solve (2.4), let us first rewrite it using joint probabilities. This will simplify treatment later on. So define $p_{ij} = p_{x_i} p_{z_j|x_i}$ to be the probability mass placed at $(x_i, z_j)$ by the joint distribution $P_{x_i} P_{z_j|x=x_i}$. Since $\sum_{j=1}^n w_{ij} = 1$

for each $i$, after dropping the inessential factor $n^{-1}$ on the objective function we can rewrite (2.4) as:

$$\max_{\{p_{ij}:i,j=1,\ldots,n\}} \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\log p_{ij} \tag{2.5}$$

$$s.t.\quad p_{ij}\geq 0,\ \sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij}=1,\ \frac{\sum_{j=1}^{n} z_j p_{ij}}{\sum_{j=1}^{n} p_{ij}}=0.$$

In a $b_n$ neighborhood of $x_i$, $w_{ij}$ assigns smaller weights to those $x_j$'s which are farther away from $x_i$. This has the effect of smoothing the empirical log likelihood at each $x_i$. The Lagrangian for this problem is given by[2]

$$\mathfrak{L}^r=\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\log p_{ij}-\mu\Big(\sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij}-1\Big)-\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i' z_j p_{ij},$$

where $\mu$ is the Lagrange multiplier for the second constraint and $\{\lambda_i\in\mathbb{R}^d : i=1,\ldots,n\}$ the set of Lagrange multipliers for the third constraint. It is easy to verify that the solution to this problem is given by

$$\hat{p}_{ij}=\frac{w_{ij}}{n+\lambda_i' z_j},$$

where each $\lambda_i$ solves

$$\sum_{j=1}^{n}\frac{w_{ij}z_j}{n+\lambda_i' z_j}=0,\quad i=1,\ldots,n. \tag{2.6}$$

The $\lambda_i$'s in (2.6) can be numerically obtained as the solution to the optimization problem

$$\min_{\varphi\in\mathbb{R}^d} -\sum_{j=1}^{n} w_{ij}\log(n+\varphi' z_j). \tag{2.7}$$

Because $\varphi\mapsto -\log(n+\varphi' z_j)$ is strictly convex, (2.7) can be uniquely solved for $\lambda_i$ in few iterations by a standard Newton-Raphson procedure. Hence we can write the restricted (i.e. under $\mathbb{E}(z|x)=0$) SEL as

$$\begin{aligned}\mathrm{SEL}^r=\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\log\hat{p}_{ij}&=\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\log\{\frac{w_{ij}}{n+\lambda_i' z_j}\}\\&=\sum_{i=1}^{n}\min_{\varphi\in\mathbb{R}^d}\sum_{j=1}^{n} w_{ij}\log\{\frac{w_{ij}}{n+\varphi' z_j}\}.\end{aligned}$$

[2]Since the objective function depends upon $p_{ij}$ only through $\log p_{ij}$, the nonnegativity constraint $p_{ij}\geq 0$ does not bind.

Next we look at the unrestricted problem, which is similar to (2.5) except that the conditional moment constraint is absent; i.e. we solve

$$\max_{\{p_{ij}:i,j=1,\ldots,n\}} \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\log p_{ij} \quad s.t. \quad p_{ij}\geq 0,\ \sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij}=1.$$

This can also be solved by the Lagrange multiplier technique to give

$$\hat{p}_{ij}=\frac{w_{ij}}{n},$$

and we can write the unrestricted SEL as

$$\mathrm{SEL}^{ur}=\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\log\{\frac{w_{ij}}{n}\}.$$

Note that for notational convenience we are ignoring the dependence of $\mathrm{SEL}^{r}$ and $\mathrm{SEL}^{ur}$ upon $n$.

An analog of the parametric likelihood ratio test statistic would then be:

$$\begin{aligned} 2(\mathrm{SEL}^{ur}-\mathrm{SEL}^{r}) &= 2\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\log(1+\frac{\lambda_i' z_j}{n}) \\ &= 2\sum_{i=1}^{n}\max_{\varphi\in\mathbb{R}^d}\sum_{j=1}^{n} w_{ij}\log(1+\frac{\varphi' z_j}{n}). \end{aligned} \tag{2.8}$$

Heuristically speaking, (2.8) will be small if the conditional moment restriction $\mathbb{E}(z|x)=0$ is indeed true. Therefore, it seems sensible to base the test for $\mathbb{E}(z|x)=0$ upon (2.8). However, as described in the next section, we will use a slightly modified version of (2.8) for our test.

Before proceeding any further, we mention some papers in the empirical likelihood literature which may be relevant to us. The basic references are, of course, the seminal papers by Owen (1988, 1990, 1991) for iid data. Using iid data Qin and Lawless (1994, 1995) look at efficiently estimating finite dimensional parameters under unconditional moment restrictions. Kitamura (1997) extends the treatment to weakly dependent data. Kitamura (1999) also describes an optimal property of empirical likelihood based tests for unconditional moment restrictions. Not much work seems to have been done as far as applying empirical likelihood to conditional moment restrictions is concerned. Some exceptions include LeBlanc and Crowley (1995), Brown and Newey (1998), and Kitamura, Tripathi, and Ahn (2000). LeBlanc and Crowley (1995) and Kitamura, Tripathi, and Ahn (2000) are mainly concerned with estimation, while Brown and Newey (1998) consider the bootstrap under a conditional moment restriction. None of these papers contain the results obtained here.

## 3. Basic Assumptions and Notation

The following basic assumptions are maintained throughout the paper:

**Assumption 3.1.** *(i)* $\{x_i, z_i\}_{i=1}^n$ *is a random sample from a probability distribution on* $S \times \mathbb{R}^d$. *(ii)* $x$ *is continuously distributed with Lebesgue density* $h : S \to \mathbb{R}$. *(iii)* $\mathbb{E}\|z\|^m < \infty$ *for some* $m > 2$. □

Notice that apart from the existence of certain moments, no other restrictions have been imposed on the distribution of $z$. For reasons about to be described, we now look at a situation where the researcher is interested in the behavior of the conditional moment $\mathbb{E}(z|x)$ on a subset $S_*$ of $S$. Therefore, consider the smoothed empirical likelihood ratio (SELR):

$$\begin{aligned}\mathrm{SELR} &= 2\sum_{i=1}^n \mathbb{I}\{x_i \in S_*\} \sum_{j=1}^n w_{ij} \log(1 + \frac{\lambda_i' z_j}{n}) \\ &= 2\sum_{i=1}^n \mathbb{I}\{x_i \in S_*\} \max_{\varphi \in \mathbb{R}^d} \sum_{j=1}^n w_{ij} \log(1 + \frac{\varphi' z_j}{n}).\end{aligned}$$

We now assume that

**Assumption 3.2.** $S_*$ *is a compact proper subset of* $S$. □

$S_*$ is equivalent to the "fixed trimming" set used in Aït-Sahalia, Bickel, and Stoker (2000) and described in Fan and Li (1996, Page 876). Paraphrasing Aït-Sahalia, Bickel, and Stoker (2000), it "... allows us to focus goodness-of-fit testing on particular ranges of the predictor variables. By choosing an appropriate $S_*$ specification tests can be tailored to the empirical question of interest." As pointed out by these authors, a consequence of this assumption is that our test will be consistent only against those alternatives which differ from the null on $S_*$. It is important to remember that the $\lambda_i$'s used in SELR are still obtained by using the entire sample of observations; i.e. the "trimming" is done after the $\lambda_i$'s have been computed.

Fixed trimming is also useful technically. As Härdle and Marron (1990, Page 66) emphasize, it allows us to avoid the usual edge effects associated with kernel estimators. For instance, and we use this many a time in the proofs, suppose we want to simplify expressions of the form $\mathbb{E}\{\mathcal{K}_{ij}\psi(x_j)|x_i\}$ where $\psi$ is some integrable function. We can use the fact that $x_i \in S_*$ to write (for small enough $b_n$)

$$\mathbb{E}\{\mathcal{K}_{ij}\psi(x_j)|x_i\} = b_n^s \int_{u \in [-1,1]^s} \mathcal{K}(u)\psi(x_i - b_n u)h(x_i - b_n u)\, du.$$

Compactness of $S_*$ is required for utilizing uniform rates of convergence for kernel estimators of conditional expectations. In addition to the previously defined symbols, the following notation is used hereafter.

**Notation.** $\mathbb{I}_i = \mathbb{I}\{x_i \in S_*\}$, $\hat{V}(x_i) = \sum_{j=1}^n w_{ij} z_j z_j'$, $V(x) = \mathbb{E}(zz'|x)$, and $V^{(lv)}$ is the $(lv)^{th}$ element of $V$. $S_\mathcal{K} = [-1,1]^s$ is the support of $\mathcal{K}$ and $\mathfrak{R}(\mathcal{K}) = \int_{S_\mathcal{K}} \mathcal{K}^2(u)\, du$ denotes the "roughness" of the kernel. The convolution of $\mathcal{K}$ with itself is given by $\mathcal{K}^*(x) = \int_{S_\mathcal{K}} \mathcal{K}(v)\mathcal{K}(x - v)\, dv$, and

$\mathcal{K}^{**} = \int_{[-2,2]^s} \{\mathcal{K}^*(u)\}^2\, du$. $\mathbb{S} = \{\xi \in \mathbb{R}^d : \|\xi\| = 1\}$ is the unit sphere in $\mathbb{R}^d$, and $vol(S_*) = \int_{S_*} dx$ is the volume of $S_*$ in $\mathbb{R}^s$. □

The next assumption imposes some additional restrictions upon the distribution of $x$.

**Assumption 3.3.** *For $1 \le l, v \le d$:*

($i$) *$h$ is bounded away from zero on $S_*$.*

($ii$) *$x \mapsto h(x)$ and $x \mapsto V^{(lv)}(x)$ are elements of $\mathrm{C}^2(S)$.*

($iii$) *$(\xi, x_i) \mapsto \xi' V(x_i)\xi$ is bounded away from zero on $\mathbb{S} \times S_*$.* □

The choice of $S_*$ is influenced by ($i$), which requires that we estimate conditional expectations in a region where we can avoid the "denominator problem". Such an assumption is regularly invoked in the kernel estimation literature to minimize complexity of mathematical details when dealing with ratios of random variables. See Newey (1994, Page 242) for a brief discussion. Although it is possible to relax ($i$) by trimming away those $x$'s in $S_*$ at which $h(x) = 0$[3], to keep the mathematical details manageable we avoid this approach. Because $\sup_{x_i \in S_*} |\frac{1}{nb_n^s}\sum_{j=1}^n \mathcal{K}_{ij} - h(x_i)| = o_p(1)$ is due to the uniform consistency of kernel density estimators and $h$ is bounded above zero on $S_*$, we can use Lemma C.1 to show that

$$\sup_{x_i \in S_*} |\{\frac{1}{nb_n^s}\sum_{j=1}^n \mathcal{K}_{ij}\}^{-1} - h^{-1}(x_i)| = o_p(1).$$

This result will be useful in the proofs. ($ii$) also ensures that

$$\sup_{(x_i,\mu) \in S_* \times [0,1]} \int_{S_\mathcal{K}} \mathcal{K}^2(u) |u' \nabla^2 \{V^{(lv)}(x_i - \mu b_n u)\, h(x_i - \mu b_n u)\} u|\, du < \infty$$

for large enough $n$. This is the remainder term in evaluating integrals of the form $\int_{u \in S_\mathcal{K}} \mathcal{K}^2(u) V(x_i - b_n u) h(x_i - b_n u)\, du$ when $V(x_i - b_n u)h(x_i - b_n u)$ is element by element expanded around $x_i$ up to second order. See, for e.g., the proof of Lemma B.4. ($iii$) implies some nice results which have been used quite a few times in the proofs. For instance, a direct implication of ($iii$) is that: ($a$) $(\xi, x_i) \mapsto \xi' \mathbb{E}\hat{V}(x_i)\xi$ is bounded away from zero on $\mathbb{S} \times S_*$ for large enough $n$, and ($b$) $\sup_{x_i \in S_*} \|V^{-1}(x_i)\| < \infty$. ($a$) is used in the proof of Lemma B.11. Since $\sup_{x_i \in S_*} \|\hat{V}(x_i) - V(x_i)\| = o_p(1)$ follows from the uniform consistency of kernel estimators, Lemma C.2 yields

$$\sup_{x_i \in S_*} \|\hat{V}^{-1}(x_i) - V^{-1}(x_i)\| = o_p(1).$$

Consequently, we can use ($b$) to show that

$$\sup_{x_i \in S_*} \|\hat{V}^{-1}(x_i)\| = O_p(1).$$

This result is used in the proof of Lemma A.1. Finally, the next assumption describes the kernel functions used to construct SELR.

[3]See, for example, Ai (1997).

**Assumption 3.4.** *The kernels $\mathcal{K}$ belong to the class of second order product kernels; i.e. for $x = (x_1, \ldots, x_s)$ let $\mathcal{K}(x) = \Pi_{i=1}^{s}\kappa(x_i)$, where $\kappa : \mathbb{R} \to \mathbb{R}$ is a symmetric, nonnegative, Lipschitz function which vanishes outside* $[-1, 1]$ *and satisfies $\int_{-1}^{1} \kappa(t)\, dt = 1$.* □

Since these kernels are employed to estimate probabilities, the use of kernels with order greater than two is ruled out. Furthermore, the nonnegativity of $\mathcal{K}$ is also explicitly used several times. See, for instance, the proof of Lemma B.1. The Lipschitz condition allows us to use the uniform convergence rates for kernel estimators obtained by Newey (1994). The bandwidth $b_n$ used in the kernels is a null sequence of positive numbers such that $nb_n^s \uparrow \infty$. Subsequently, additional restrictions will be imposed upon the choice of $b_n$.

## 4. The Test Statistics and their Distributions under the Null

As mentioned earlier, the test for $\mathbb{E}(z|x) = 0$ will be based on SELR. The first step is to transform SELR so that we can apply a CLT due to de Jong (1987). Following Lemma A.1 we can write

$$\mathrm{SELR} = \mathrm{T}_n + O_p(\{\frac{\log n}{n^{\frac{1}{2}-\frac{1}{m}} b_n^s}\}^2) + O_p(\{\frac{\log n}{n^{\frac{1}{3}-\frac{2}{m}} b_n^s}\}^{3/2}), \quad \text{where}$$

$$\mathrm{T}_n = \sum_{i=1}^{n} \mathbb{I}_i (\sum_{j=1}^{n} w_{ij} z_j') \hat{V}^{-1}(x_i) (\sum_{j=1}^{n} w_{ij} z_j).$$

Now use the summation identity in (D.1) to decompose $\mathrm{T}_n = \mathrm{T}_{n,1} + \mathrm{T}_{n,2} + \mathrm{T}_{n,3} + \mathrm{T}_{n,4} + \mathrm{T}_{n,5}$, where

$$\mathrm{T}_{n,1} = \mathcal{K}^2(0) \sum_{i=1}^{n} \mathbb{I}_i \frac{z_i' \hat{V}^{-1}(x_i) z_i}{\{\sum_{u=1}^{n} \mathcal{K}_{iu}\}^2},$$

$$\mathrm{T}_{n,2} = \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \mathbb{I}_i w_{ij}^2 z_j' \hat{V}^{-1}(x_i) z_j,$$

$$\mathrm{T}_{n,3} = \mathcal{K}(0) \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \mathbb{I}_i \frac{z_i' \hat{V}^{-1}(x_i) z_j w_{ij}}{\sum_{u=1}^{n} \mathcal{K}_{iu}},$$

$$\mathrm{T}_{n,4} = \mathcal{K}(0) \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \mathbb{I}_i \frac{w_{ij} z_j' \hat{V}^{-1}(x_i) z_i}{\sum_{u=1}^{n} \mathcal{K}_{iu}},$$

$$\mathrm{T}_{n,5} = \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \sum_{t=1, t\neq j\neq i}^{n} \mathbb{I}_i w_{ij} z_j' \hat{V}^{-1}(x_i) z_t w_{it}.$$

The asymptotic behavior of these terms under $\mathrm{H}_0$ can be obtained from Lemmas A.2 – A.6. These results are summarized below:

$$\mathrm{T}_{n,1} \overset{\text{Lemma A.2}}{=} O_p(\frac{1}{nb_n^{2s}}),$$

$$\mathrm{T}_{n,2} \overset{\text{Lemma A.3}}{=} b_n^{-s}\{d\,\mathfrak{R}(\mathcal{K})\,vol(S_*) + O_p(\sqrt{\frac{\log n}{nb_n^s}} + b_n^2)\},$$

$$\mathrm{T}_{n,3} \overset{\text{Lemma A.4}}{=} o_p(\frac{1}{\sqrt{nb_n^{2s}}}), \quad \mathrm{T}_{n,4} \overset{\text{Lemma A.5}}{=} o_p(\frac{1}{\sqrt{nb_n^{2s}}}),$$

$$b_n^{s/2}\mathrm{T}_{n,5} \xrightarrow[\text{Lemma A.6}]{d} \mathrm{N}(0,\sigma^2), \qquad \sigma^2 = 2d\mathcal{K}^{**}\,vol(S_*).$$

Although $b_n^{s/2}\mathrm{T}_{n,5}$ is asymptotically distributed as $\mathrm{N}(0,\sigma^2)$ and

$$b_n^{s/2}\mathrm{T}_{n,1} = o_p(1),\ b_n^{s/2}\mathrm{T}_{n,3} = o_p(1),\ b_n^{s/2}\mathrm{T}_{n,4} = o_p(1) \text{ as } nb_n^{3s/2} \uparrow \infty,$$

$b_n^{s/2}\mathrm{T}_{n,2}$ explodes as $n \uparrow \infty$. Therefore, we have to center $\mathrm{T}_n$ appropriately if we want a test statistic with a valid asymptotic distribution. We do this by subtracting[4] the troublesome quantity $b_n^{s/2}\mathrm{T}_{n,2}$ from SELR.

Let

$$\zeta_{1,n} \overset{\text{def}}{=} \{b_n^{s/2}\mathrm{SELR} - b_n^{s/2}\mathrm{T}_{n,2}\}/\sigma$$

denote the first version of our statistic for testing $\mathbb{E}(z|x) = 0$. Obtaining the asymptotic distribution of $\zeta_{1,n}$ is straightforward. First, it is easily seen that

$$\zeta_{1,n} = \{b_n^{s/2}\mathrm{T}_{n,1} + b_n^{s/2}\mathrm{T}_{n,3} + b_n^{s/2}\mathrm{T}_{n,4} + b_n^{s/2}\mathrm{T}_{n,5}\}/\sigma \\ + O_p(\{\frac{\log n}{n^{\frac{1}{2}-\frac{1}{m}}b_n^{3s/4}}\}^2) + O_p(\{\frac{\log n}{n^{\frac{1}{3}-\frac{2}{m}}b_n^{2s/3}}\}^{3/2}).$$

Then using Lemmas A.2–A.6, the following result is almost immediate.

**Theorem 4.1.** *Let* $\mathbb{E}\|z_1\|^m < \infty$ *for some* $m > 6$ *and choose* $b_n = n^{-\alpha}$ *for* $0 < \alpha < \frac{1}{2s}(1 - \frac{6}{m})$. *Then under* $\mathrm{H}_0$, $\zeta_{1,n} \xrightarrow{d} \mathrm{N}(0,1)$.

The test for $\mathrm{H}_0$ can be implemented by comparing $\zeta_{1,n}$ with critical values obtained from a standard normal distribution. Namely, to obtain a one sided size-$\gamma$ test we reject $\mathrm{H}_0$ if $\zeta_{1,n} > z_\gamma$, where $z_\gamma$ denotes the $\gamma$-cutoff point for the standard normal distribution; i.e. $\Pr\{\mathrm{N}(0,1) \geq z_\gamma\} = \gamma$. Notice that $\sigma^2$ does not depend upon any unknown parameters and can be calculated analytically.

Although our first test is simple to implement and does not require the estimation of any variance term, we do have to calculate $\mathrm{T}_{n,2}$ to obtain $\zeta_{1,n}$.

[4]Note that subtracting $b_n^{s/2}\mathrm{T}_{n,2}$ does not lead to any loss of information as far as testing $\mathbb{E}(z|x) = 0$ is concerned. To verify this, look at Lemmas A.3 and A.9 which show that the asymptotic behavior of $\mathrm{T}_{n,2}$ remains unchanged under $\mathrm{H}_0$ and the sequence of local alternatives $\mathrm{H}_{1n}$ defined in the next section; i.e. $\mathrm{T}_{n,2}$ is asymptotically uninformative about the null hypothesis.

Even this calculation can be eliminated for the special case $s < 4$; i.e. when we have at most three explanatory variables. To see this, from Lemma A.3 observe that

$$b_n^{s/2}\mathrm{T}_{n,2} = b_n^{-s/2} d\,\mathfrak{R}(\mathcal{K})\,vol(S_*) + O_p\Big(\sqrt{\frac{\log n}{nb_n^{2s}}} + b_n^{2-\frac{s}{2}}\Big).$$

But the choice of $b_n$ in Theorem 4.1 implies that $\frac{\log n}{nb_n^{2s}} \downarrow 0$. Therefore, when $s < 4$ we propose to use the second version of our statistic defined as

$$\zeta_{2,n} = \{b_n^{s/2}\mathrm{SELR} - b_n^{-s/2} d\,\mathfrak{R}(\mathcal{K})\,vol(S_*)\}/\sigma.$$

The following corollary is obvious.

**Corollary 4.1.** *Let* $s < 4$, $\mathbb{E}\|z_1\|^m < \infty$ *for some* $m > 6$, *and* $b_n = n^{-\alpha}$ *for* $0 < \alpha < \frac{1}{2s}(1 - \frac{6}{m})$. *Then under* $\mathrm{H}_0$, $\zeta_{2,n} \xrightarrow{d} \mathrm{N}(0, 1)$.

In practice, $\zeta_{2,n}$ seems more useful than $\zeta_{1,n}$ because $s < 4$ is a reasonable bound for most applications of nonparametric regression.

A nice interpretation of Corollary 4.1 can be obtained by observing that we can express its result as

$$\frac{\mathrm{SELR} - c_1\gamma_n}{c_2\sqrt{2\gamma_n}} \xrightarrow{d} \mathrm{N}(0, 1), \tag{4.1}$$

where $c_1 = \mathfrak{R}(\mathcal{K})$, $c_2 = \sqrt{\mathcal{K}^{**}}$, and $\gamma_n = b_n^{-s}\,d\,vol(S_*)$. This can be regarded as an analog of Wilks's theorem: If SELR were distributed as a $\chi^2$ random variable with $c_1\gamma_n$ degrees of freedom, and $\mathfrak{R}(\mathcal{K}) = \mathcal{K}^{**}$ so that $c_2 = \sqrt{c_1}$, then (4.1) can be interpreted as the normal approximation of a $\chi^2$ random variable with large degrees of freedom.

## 5. Distribution of $\zeta_{1,n}$ under Local Alternatives

In this section we obtain an expression for the power of the test under a sequence of local alternatives. Assume that

**Assumption 5.1.** $\delta = (\delta^{(1)}, \ldots, \delta^{(d)})$ *is a vector of* $\mathrm{L}^2(S)$ *functions which are continuous on* $S$. □

Square integrability of each component ensures that $\delta^{(l)}$ is bounded in probability for $1 \le l \le d$. Continuity of $x_i \mapsto \delta^{(l)}(x_i)$ is used, for instance in the proofs of Lemma A.7 and Lemma A.12, to bound $g_n(x_i) = \int_{S_\mathcal{K}} \mathcal{K}(u)\delta^{(l)}(x_i - b_n u)h(x_i - b_n u)\,du$ when $x_i \in S_*$. To see this, observe that we can write

$$|g_n(x_i)| \le \sup_{u\in S_\mathcal{K}} |\delta^{(l)}(x_i - b_n u)| \sup_{u\in S_\mathcal{K}} h(x_i - b_n u) \int_{S_\mathcal{K}} \mathcal{K}(u)\,du.$$

But continuity of $\delta^{(l)}$ and $h$ implies that the maps $x_i \mapsto \sup_{u\in S_\mathcal{K}} |\delta^{(l)}(x_i - b_n u)|$ and $x_i \mapsto \sup_{u\in S_\mathcal{K}} h(x_i - b_n u)$ are continuous on $S$ for each $n$. Therefore, since $S_*$ is a compact subset of $S$ and $\mathcal{K}$ integrates to one on $S_\mathcal{K}$, $g_n(x_i)$ is uniformly bounded on $S_*$ for each $n$. Furthermore, when $x_i \in S_*$ this argument also shows that we can apply Lebesgue's dominated convergence theorem to obtain $g_n(x_i) = \delta^{(l)}(x_i)h(x_i) + o(1)$.

Henceforth, let $\{x_i, \tilde{z}_i\}_{i=1}^n$ denote a collection of iid random vectors satisfying Assumption 3.1 such that $\mathbb{E}(\tilde{z}_i|x_i) = 0$ and $\mathbb{E}(\tilde{z}_i\tilde{z}_i'|x_i) = V(x_i)$. The $\tilde{z}_i$'s are unobserved but we do observe $z_i$, which is distributed according to the sequence of local alternatives

$$\mathrm{H}_{1n} : z_i = \tilde{z}_i + \frac{\delta(x_i)}{n^{1/2}b_n^{s/4}} \qquad i = 1, \dots, n.$$

Notice that $\mathbb{E}(z_i|x_i) = \frac{\delta(x_i)}{n^{1/2}b_n^{s/4}}$ under $\mathrm{H}_{1n}$. As Lemma A.7 shows, the result of Lemma A.1 remains valid under $\mathrm{H}_{1n}$. Defining $\zeta_{1,n} = \{b_n^{s/2}\mathrm{SELR} - b_n^{s/2}\mathrm{T}_{n,2}\}/\sigma$ as before, we can thus write

$$\zeta_{1,n} \overset{\mathrm{H}_{1n}}{=} \{b_n^{s/2}\mathrm{T}_{n,1} + b_n^{s/2}\mathrm{T}_{n,3} + b_n^{s/2}\mathrm{T}_{n,4} + b_n^{s/2}\mathrm{T}_{n,5}\}/\sigma + O_p(\{\frac{\log n}{n^{\frac{1}{2}-\frac{1}{m}}b_n^{3s/4}}\}^2) + O_p(\{\frac{\log n}{n^{\frac{1}{3}-\frac{2}{m}}b_n^{2s/3}}\}^{3/2}).$$

Using Lemmas A.8–A.12 it is now easy to obtain the following result.

**Theorem 5.1.** *Let* $\mathbb{E}\|\tilde{z}_1\|^m < \infty$ *for some* $m > 6$ *and choose* $b_n = n^{-\alpha}$ *for* $0 < \alpha < \frac{1}{2s}(1 - \frac{6}{m})$. *Also let* $\mu = \mathbb{E}[\mathbb{I}\{x_1 \in S_*\}\delta'(x_1)V^{-1}(x_1)\delta(x_1)]$. *Then under* $\mathrm{H}_{1n}$, $\zeta_{1,n} \xrightarrow{d} \mathrm{N}(\mu/\sigma, 1)$.

Therefore, $\Pr\{\zeta_{1,n} > z_\gamma\} \xrightarrow{\mathrm{H}_{1n}} 1 - \Phi(z_\gamma - \frac{\mu}{\sigma})$, where $\Phi$ denotes the cdf of a $\mathrm{N}(0,1)$ random variable. The same result holds for $\zeta_{2,n}$ when $s < 4$.

## 6. Asymptotic Optimality Of The SELR Test

As noted in the introduction, there are alternative tests for conditional moment restrictions available in the literature. All of these tests are nonparametric and are consistent against general alternatives. There is, of course, a price one pays for this generality: nonparametric tests tend to have lower power than parametric ones. Therefore, it is important to find a nonparametric test with good power properties.

This section identifies an optimal test among a class of conditional moment restrictions tests. Aït-Sahalia, Bickel, and Stoker (2000) provide a convenient framework for this purpose. They consider a testing procedure based on a weighted sum of squared residuals from kernel regression. Many earlier tests, at least asymptotically, can be regarded as a special case of this test with a particular choice of weighting function. Härdle and Mammen (1993), Fan and Li (1996), Zheng (1996), and our SELR test, for example, fall into this category. Hong and White (1995) apply a similar principle, though they use series instead of kernels.

To simplify our argument let $d = 1$, $s = 1$, and $S_* = [0, 1]$. In implementing the Aït-Sahalia, Bickel, and Stoker test the researcher chooses a piecewise smooth, bounded, and square integrable weight function $a : [0,1] \to \mathbb{R}_+$ and calculates

$$G(a) = b_n \sum_{i=1}^{n} \hat{\mathbb{E}}^2(z|x_i)a(x_i).$$

The statistic for testing $\mathrm{H}_0$ proposed by Aït-Sahalia et al. is

$$\tau(a) = \frac{b_n^{-1/2}\{G(a) - \mathfrak{R}(\mathcal{K})\int_0^1 V(x)a(x)dx\}}{\sqrt{2\mathcal{K}^{**}\int_0^1 V^2(x)a^2(x)\,dx}}. \tag{6.1}$$

We can replace $V(x)$ with an appropriate consistent estimator without affecting the asymptotic properties of the test. Since $\tau(ca) = \tau(a)$ for any $c \neq 0$, w.l.o.g we assume that $\int_0^1 a^2(x)dx = 1$. Now let

$$M(a,\delta) = \frac{\int_0^1 \delta^2(x)a(x)h(x)\,dx}{\sqrt{2\mathcal{K}^{**}\int_0^1 V^2(x)a^2(x)\,dx}}. \tag{6.2}$$

As Aït-Sahalia et al. show, under $\mathrm{H}_{1n}$

$$\tau(a) \xrightarrow{d} \mathrm{N}(M(a,\delta),1). \tag{6.3}$$

The asymptotic power of their test with critical value $z_\gamma$ is thus given by

$$\pi(a,\delta) = 1 - \Phi(z_\gamma - M(a,\delta)). \tag{6.4}$$

Comparing (6.4) and Theorem 5.1, we can see that our SELR test is asymptotically equivalent to the $\tau(a)$ test with the weighting scheme

$$a_{\mathrm{SELR}}(x) = \frac{1}{V(x)\sqrt{\int_0^1 V^{-2}(x)\,dx}}. \tag{6.5}$$

We shall demonstrate that this choice of weighting, which is implicitly achieved by the SELR test, is optimal in a certain sense.

If $\delta$ is known counterfactually, it is easy to derive the optimal weighting function that maximizes (6.2). For a known $\delta$, an application of the Cauchy-Schwarz inequality on (6.2) shows that (6.4) is maximized by choosing

$$a(x,\delta) = \frac{\delta^2(x)h(x)}{V^2(x)\sqrt{\int_0^1 \delta^4(x)V^{-4}(x)h^2(x)\,dx}}. \tag{6.6}$$

The notation $a(x,\delta)$ indicates that the optimal choice of $a$ depends on $\delta$. This result is not terribly useful since $\delta$ is unknown in practice. It is also clear from (6.6) that there is no uniformly (in $\delta$) optimal test. This resembles the multiparameter optimal testing problem considered in the seminal paper of Wald (1943).

Wald shows that the likelihood ratio test, and other asymptotically equivalent tests, for a hypothesis about finite dimensional parameters is optimal in terms of an average power criterion. Loosely put, he considers a weighted average of the power function where uniform weights are given along each probability contour of the distribution of the estimator he uses (mle). This criterion is natural and attractive since it is impartial – it puts heavy (light) weights in directions where the detection of departures from the null is difficult (easy). This approach has been used in the literature quite effectively. For example, Andrews and Ploberger (1994) consider optimal inference in a nonstandard testing problem. They derive a test that is optimal with respect to a Wald-type average power criterion. Their optimal test performs well in finite samples (see Andrews and Ploberger, 1996) indicating the practical relevance of Wald's approach.

Our testing problem is different from the ones considered by Wald in that instead of being finite dimensional, our parameter of interest is an unknown function. A natural extension of Wald's approach is to consider a probability measure on an appropriate space of functions and let the measure mimic the distribution of the "estimator." Then the local average power criterion is obtained by integrating (6.4) against the probability measure. Note that the tests we are comparing rely on the kernel regression estimator $\hat{\mathbb{E}}(z|x)$, either explicitly or implicitly. Therefore, we propose to use a probability measure that approximates the distribution of the sample path of $\hat{\mathbb{E}}(z|x)$.

So let $\tilde{\delta}$ be a C([0, 1])-valued random variable given by

(6.7)

$$\tilde{\delta}(x) = V^{1/2}(x)\, h^{-1/2}(x)\, y(x) \quad \text{and} \quad y(x) = \int_0^1 k(\frac{x}{\beta} - z)\, dW(z - \lfloor z \rfloor),$$

where $W$ is the standard Brownian motion on $[0, 1]$, $k(\cdot)$ an appropriate weighting function, $\beta$ a positive adjustable parameter and $\lfloor z \rfloor$ the integer part of $z$. For each $x$ in $[0, 1]$, $y(x)$ is a stochastic integral. Note the use of $dW(z - \lfloor z \rfloor)$ as the integrator. This implies that the covariance kernel $r(s) = \mathbb{E}[y(x)y(x + s)]$ of the Gaussian process $y$ is circular; i.e. $r(s) = r(1-s)$. Circular processes are widely used for analyzing stationary processes on a finite interval (see, for example, Hannan, 1970 and Priestley, 1981). In our case it lets us avoid treating $y(x)$'s close to the end points of the interval $[0, 1]$ differently from the ones in the middle. Consequently, for an arbitrary function $f$ such that the integral $\int_0^1 f(y(x))\, dx$ is well defined, the joint distribution of the bivariate random vector $(\int_0^1 f(y(x))\, dx, y(x_0))$ does not depend on the location $x_0 \in [0, 1]$. Other properties of $\tilde{\delta}$, such as its Gaussianity, are not important in our argument below.

Note that the variance function of $\tilde{\delta}(x)$ coincides with the asymptotic variance function of $\hat{\mathbb{E}}(z|x)$ up to scale. This is one of the features we intend to replicate by using $\tilde{\delta}$. The Gaussian process $\tilde{\delta}$ is constructed based on an approximation of $\hat{\mathbb{E}}(z|x)$ derived by Liero (1982). Also see Johnston (1982) and Härdle (1989) for related results. In our theory however, $k$ and $\beta$ do not have to be the same as $\mathcal{K}$ and $b_n$. $k$ determines the pattern of autocorrelations of $y(x)$ and $\beta$ is used for scaling $x$. A large $\beta$ and a spread-out $k$ correspond to stronger dependence, yielding paths of $y$ and $\tilde{\delta}$ that look smoother. Our optimality result does not depend on the choice of $\beta$ and $k$.

We are now ready to define our average power concept. Let $Q$ be the probability measure induced by $\tilde{\delta}$ on C([0, 1]). Using (6.7) rewrite the random variable $M(a, \tilde{\delta})$ as

$$M(a,\tilde{\delta}) = \frac{\int_0^1 V(x)y^2(x)a(x)\,dx}{\sqrt{2\mathcal{K}^{**}\int_0^1 V^2(x)a^2(x)\,dx}} = \frac{1}{\sqrt{2\mathcal{K}^{**}}}\int_0^1 A(x)y^2(x)\,dx,$$

where

$$A(x) = \frac{V(x)a(x)}{\sqrt{\int_0^1 V^2(x)a^2(x)\,dx}}. \tag{6.8}$$

$\int_0^1 A^2(x)\,dx = 1$ and it is sometimes convenient to deal with $A$ rather than $a$. Note that $M(a,\tilde{\delta}) = M(A/V,\tilde{\delta})$. Let $F_A$ be the cdf of $M(A/V,\tilde{\delta})$. The average asymptotic power of the test proposed by Aït-Sahalia et al. (see (6.4)) is the following functional of $A$:

$$\bar{\pi}(A) = \int \pi(A/V,\tilde{\delta})dQ(\tilde{\delta}) = \int_0^\infty [1-\Phi(z_\gamma - m)]\,F_A(dm). \tag{6.9}$$

Observe that the integrand in (6.9) is strictly increasing in $m$. So if there exists a piecewise smooth, bounded, square integrable function $A^* : [0,1] \mapsto \mathbb{R}_+$ such that $\int_0^1 A^{*2}(x)\,dx = 1$ and for all $A$ the cdf $F_{A^*}$ first-order stochastically dominates[5] $F_A$, then $A^*$ maximizes $\bar{\pi}(A)$. By (6.8), the optimal weighting function $a^*$ is given by

$$a^*(x) = \frac{A^*(x)}{V(x)\sqrt{\int A^{*2}(x)/V^2(x)\,dx}}.$$

To find $A^*$, fix $m \in \mathbb{R}$ arbitrarily and consider solving the following variational problem over all piecewise smooth, bounded, square integrable functions from $[0,1] \to \mathbb{R}_+$:

$$\min_A F_A(m) \quad s.t. \quad \int_0^1 A^2(x)dx = 1. \tag{6.10}$$

For any $x_0 \in [0,1]$ let $F_A(m|y(x_0))$ be the conditional cdf of $M(A/V,\tilde{\delta})$ given $y(x_0)$. $f_A(m|y(x_0))$ denotes the conditional pdf corresponding to $F_A(m|y(x_0))$. Now it is clear that

$$F_A(m) = \mathbb{E}_{y(x_0)}\left[F_A(m|y(x_0))\right],$$

where the symbol $\mathbb{E}_{y(x_0)}$ indicates that the expectation is over $y(x_0)$. Furthermore,

$$\begin{aligned}\frac{\partial F_A(m|y(x_0))}{\partial A(x_0)} &= \frac{\partial \mathbb{E}\left[\mathbb{I}\{\int_0^1 A(x)y^2(x)dx < m\}|y(x_0)\right]}{\partial A(x_0)} \\ &= y^2(x_0)f_A(m|y(x_0)).\end{aligned}$$

[5] i.e. $F_A(m) \geq F_{A^*}(m)$ for all $m$.

These results imply that

$$\frac{\partial F_A(m)}{\partial A(x_0)} = \mathbb{E}_{y(x_0)}\left[y^2(x_0) f_A(m|y(x_0))\right] \qquad \text{for all } x_0 \in [0,1].$$

Thus the Euler-Lagrange equation for the variational problem (6.10) is

$$\mathbb{E}_{y(x_0)}\left[y^2(x) f_{A^*}(m|y(x_0))\right] = 2\lambda A^*(x_0) \quad \text{for all } x_0 \in [0,1], \tag{6.11}$$

where $\lambda$ is the Lagrange multiplier for the constraint in (6.10) and $A^*$ the solution. To solve (6.11) we use a guess and verify approach. So suppose that $A^*(x) = \mathbb{I}\{x \in [0,1]\}$. Clearly, this is a feasible guess. As noted in our earlier discussion on the nature of the random process $y$, the joint distribution of $M(A^*/V, \tilde{\delta}) = \frac{1}{\sqrt{2\mathcal{K}^{**}}} \int_0^1 y^2(x)dx$ and $y(x_0)$ does not depend on $x_0 \in [0,1]$. Therefore,

$$\mathbb{E}_{y(x_0)}\left[y^2(x_0) f_{A^*}(m|y(x_0))\right] \stackrel{\text{def}}{=} K \text{ (say)}$$

does not depend on $x_0 \in [0,1]$. So (6.11) is satisfied with $A^*(x) = \mathbb{I}\{x \in [0,1]\}$ and $\lambda = K/2$. We have verified that $A^*(x) = \mathbb{I}\{x \in [0,1]\}$ solves (6.10). The optimal $a$ corresponding to $A^*(x) = \mathbb{I}\{x \in [0,1]\}$ is

$$a^*(x) = \frac{\mathbb{I}\{x \in [0,1]\}}{V(x)\sqrt{\int_0^1 V^{-2}(x)\, dx}}.$$

Comparing this with (6.5), we immediately obtain that the weight $a_{\text{SELR}}$ is optimal.

The above result shows that the SELR test attains the maximum average local power. An alternative way of achieving this optimality is to estimate $a^*$ by

$$\hat{a}^*(x) = \frac{\mathbb{I}\{x \in [0,1]\}}{\hat{V}(x)\sqrt{\int_0^1 \hat{V}^{-2}(x)\, dx}},$$

where $\hat{V}(x) = \sum_{j=1}^n z_j^2 \mathcal{K}(\frac{x-x_j}{b_n}) / \sum_{j=1}^n \mathcal{K}(\frac{x-x_j}{b_n})$. We then use $\hat{a}^*$ to calculate $G$ for the test statistic in (6.1). While this approach is valid asymptotically, such a "plug-in" method often leads to poor finite sample behavior. At the very least it would require a good nonparametric estimator of $V(x)$. An advantage of our statistic over plug-in statistics is that this optimal weighting is carried out automatically and implicitly, eliminating the need of estimating $V(x)$. This feature is similar to the "internal studentization" property of other empirical likelihood ratio statistics emphasized in the literature. Empirical evidence suggests that internal studentization often improves finite sample properties of the tests substantially. See, for example, Fisher, Hall, Jing, and Wood (1996).

## 7. Simulation Results

In this section we compare the SELR test with two other tests mentioned in the introduction, namely, the tests proposed by Aït-Sahalia, Bickel, and

Stoker (2000) and Zheng (1996). From (6.1) it is easy to see that Zheng's statistic is asymptotically equivalent to the one proposed by Aït-Sahalia et al. under the weighting scheme $a(x) = h(x)$.

Our experimental design is as follows. A random sample $\{x_i, z_i\}_{i=1}^n$ is drawn from the following conditionally heteroscedastic model:

$$z_i = d(x_i) + V^{1/2}(x_i)\varepsilon_i, \quad x_i \overset{\text{iid}}{\sim} \text{Uniform}[0,1], \quad \varepsilon_i \overset{\text{iid}}{\sim} \text{N}(0,1).$$

The three tests are applied to the simulated data to detect the deviation of the function $d$ from zero. Obviously $S = [0,1]$ under this specification.

The SELR test is implemented with the trimming set $S_* = [0.05, 0.95]$ and the statistic $\zeta_{2,n}$. Using $\zeta_{1,n}$ did not appear to change our simulation results significantly. The Aït-Sahalia et al. test is also carried out using the same trimming set and a weighting function that is constant over it; namely, we use (6.1) with the weighting function $a(x) = \mathbb{I}\{x \in [0.05, 0.95]\}$. Since the $x$'s are uniformly distributed in our experimental design, the uniformly weighted Aït-Sahalia et al. test and the Zheng test (hereafter designated as "$\text{ABS}_{\text{uw}}$" and "Zheng") are equivalent in terms of their local power functions. All of the three tests are implemented using a Gaussian kernel with bandwidth $b_n = 0.5n^{-1/4.25}\hat{\sigma}_x$, where $\hat{\sigma}_x$ is the sample standard deviation of $x_1, \ldots, x_n$. A similar bandwidth sequence is used by Aït-Sahalia et al. in their simulation study.

It should be noted that Zheng assumes that the conditioning variable $x$ has a distribution with unbounded support. Our simulation design violates this condition. Although it is certainly possible to address this problem by suitably modifying Zheng's statistic, we implement Zheng's test as suggested in his paper. Therefore, the reader should interpret our simulation results for Zheng's test with some caution.

Three specifications of the heteroscedasticity function are considered:

$$V(x) = \begin{cases} 1 + x^2 & \text{in Tables 1 and 2,} \\ x & \text{in Tables 3 and 4,} \\ 0.5 + \mathbb{I}\{x \geq 0.5\} & \text{in Tables 5 and 6.} \end{cases}$$

For each specification of $V(x)$, two choices of $d$ are used:

$$d(x) = \begin{cases} d_{V^{1/2}}(x) & \text{in Tables 1, 3 and 5} \\ d_V(x) & \text{in Tables 2, 4 and 6,} \end{cases}$$

where

$$d_{V^{1/2}}(x) = c\,\mathbb{I}\{x \in S_*\}\,V^{1/2}(x) \quad \text{and} \quad d_V(x) = c\,\mathbb{I}\{x \in S_*\}\,V(x).$$

In each table the constant $c$ varies over the set $\{0.0, 0.1, ..., 0.5\}$.

The above specification is motivated by the following considerations. Recall that (6.6) implies the optimal weighting function for a given (and unknown) $\delta$. For example, if $\delta = d_{V^{1/2}}$ the optimal weighting function is proportional to $\mathbb{I}\{x \in S_*\}V^{-1}(x)$, which the SELR test implicitly achieves. Likewise, if $\delta = d_V$ the optimal weighting function is a constant and thus

$\mathrm{ABS_{uw}}$ and Zheng are optimal. In short, according to our asymptotic theory, $d_{V^{1/2}}$ favors the SELR test and $d_V$ the other two.

Moreover, $d_{V^{1/2}}$ favors SELR to the same degree as $d_V$ does $\mathrm{ABS_{uw}}$ and Zheng in the following sense: Recall the definition of the asymptotic mean functional $M(a, \delta)$ in (6.2). Since asymptotically $a = \mathbb{I}\{x \in S_*\}V^{-1}$ in SELR and $a = \mathbb{I}\{x \in S_*\}$ in the other two tests,

$$R(\delta) = M(\mathbb{I}\{x \in S_*\}V^{-1}, \delta)/M(\mathbb{I}\{x \in S_*\}, \delta)$$

gives the ratio of the asymptotic means of the two types of test statistics as a functional of the local alternative $\delta$. If $R(\delta)$ is substantially larger than unity, the local power of SELR exceeds the local power of $\mathrm{ABS_{uw}}$ and Zheng by a large margin. It is easy to see that $R(d_{V^{1/2}})$ is the reciprocal of $R(d_V)$. In this sense we treat the two types of tests symmetrically by considering $d_{V^{1/2}}$ and $d_V$.

Tables 1-6 (see Appendix E) report simulated rejection probabilities of the three tests calculated from 1000 Monte Carlo replications. The size of the three tests, shown in rows with $c = 0$, appears to be reasonable although all of the tests exhibit some size distortion. The subsequent discussion therefore focuses on size-corrected power. The size correction is implemented by using distributions of $10,000$ draws of each statistic under the null hypothesis[6].

Table 1 shows the results with $V(x) = 1 + x^2$ and $d(x) = d_{V^{1/2}}(x)$. Though the SELR test tends to have somewhat higher power than the other two, the differences are marginal. The rejection frequencies with $n = 250$ are considerably higher than those with $n = 100$, in accordance with the consistency property of the three tests. With $c = 0.5$ and $n = 250$ all the tests are always able to reject the alternative.

Table 2 displays the results for the same $V(x)$ but with $d(x) = d_V(x)$, which favors $\mathrm{ABS_{uw}}$ (and Zheng as well, although the boundedness of $S$ may affect its performance as noted earlier). The power of $\mathrm{ABS_{uw}}$ sometimes exceeds that of SELR as the asymptotic theory suggests, though the differences tend to be quite small.

Tables 3-6 provide a clearer picture. In Table 3, where the SELR test is asymptotically optimal, it actually tends to have substantially higher power than the other two. When we choose an alternative that favors $\mathrm{ABS_{uw}}$ (i.e. Table 4), it has power that is only marginally higher than SELR's when $n = 250$ and the ranking is often reversed when $n = 100$. SELR also performs well in Tables 5 and 6, where $V(x) = 0.5 + \mathbb{I}\{x \geq 0.5\}$. SELR tends to be considerably more powerful than the other two tests in Table 5 where the alternative favors SELR, while the results are rather mixed in Table 6 where at least $\mathrm{ABS_{uw}}$ should perform well asymptotically.

While the scope of the simulation experiment is rather limited, the following picture emerges: $(i)$ All the tests have satisfactory size and power properties, though some size distortion remains even for $n = 250$; $(ii)$ While

[6]Note that in the tables the size corrected power for $c = 0$ does not match nominal size due to simulation errors.

asymptotic theory predicts qualitative features of the local power of the tests reasonably well, SELR seems to have better finite sample power properties than the other tests. Namely, SELR tends to have considerably higher power when the asymptotic theory predicts so, and keeps up with the other tests quite well even when the asymptotic theory does not favor it.

## 8. Conclusion

The results obtained so far are quite encouraging. The SELR test is easy to construct and straightforward to implement. It is asymptotically normal under the null hypothesis and is able to detect local alternatives which converge to the null at rate $n^{-1/2}b_n^{-s/4}$. In large samples it also possesses an optimality property.

As far as extending this work is concerned, testing conditional moment restrictions of the form $\mathbb{E}\{g(z,\theta_0)|x\} = 0$ is important. The challenge here is to obtain the asymptotic distribution of the test statistic when $\theta_0$ is replaced by a consistent estimator (say $\hat{\theta}$), although one would expect that the parametric rate of convergence of $\hat{\theta}$ should leave the asymptotic distribution unchanged. In the future we also intend to do some work on choosing a data driven bandwidth. This is important as the choice of bandwidth influences the optimization problem in (2.5): If $b_n$ is too small, say $b_n = 0$ in the extreme case so that $w_{ij} = 1$ if $i = j$ and zero otherwise, then (2.5) reduces to maximizing (2.1) under (2.2) and our procedure breaks down. On the other hand if $b_n$ is too large, say $b_n = \infty$ in the extreme case so that $w_{ij} = 1/n$, then (2.5) imposes the weaker restriction $\mathbb{E}z = 0$ instead of the stronger restriction $\mathbb{E}(z|x) = 0$. Therefore, one has to be careful when picking a bandwidth to implement the test. □

## Appendix A. Proofs of Main Results

**Notation.** The following symbols are used throughout the proofs: The letter $c$ denotes a generic constant which may differ from case to case, $\mathcal{X} = \{x_1, \dots, x_n\}$, $\hat{h}(x_i) = \frac{1}{nb_n^s}\sum_{u=1}^n \mathcal{K}_{iu}$, $\hat{H}(x_i) = \hat{V}(x_i)\hat{h}^2(x_i)$, $H(x_i) = V(x_i)h^2(x_i)$, $\hat{\Omega}(x_i) = \frac{1}{nb_n^s}\sum_{j=1}^n \mathcal{K}_{ij}z_j z_j' = \hat{V}(x_i)\hat{h}(x_i)$, and $I_* = \{1 \le i \le n : x_i \in S_*\}$. Following the discussion after Assumption 3.3, it is straightforward to show that $\hat{H}^{-1}(x_i)$ is element by element uniformly bounded in probability on $S_*$ for large enough $n$. Similarly, $H^{-1}(x_i)$ is element by element uniformly bounded on $S_*$. These facts will be used subsequently. $I_{d\times d}$ denotes the $d \times d$ identity matrix, $O_{d\times d}$ the $d \times d$ null matrix, and "elt. by elt." is shorthand for "element by element." □

**Lemma A.1.** *Let $\mathbb{E}\|z_1\|^m < \infty$ for some $m > 2$ and choose $b_n = n^{-\alpha}$ for $0 < \alpha < \frac{1}{s}(1 - \frac{2}{m})$. Then under* $\mathrm{H}_0$*, we can write*

$$\mathrm{SELR} = \mathrm{T}_n + O_p(\{\frac{\log n}{n^{\frac{1}{2}-\frac{1}{m}}b_n^s}\}^2) + O_p(\{\frac{\log n}{n^{\frac{1}{3}-\frac{2}{m}}b_n^s}\}^{3/2}),$$

*where* $\mathrm{T}_n = \sum_{i=1}^{n} \mathbb{I}_i (\sum_{j=1}^{n} w_{ij} z_j') \hat{V}^{-1}(x_i) (\sum_{j=1}^{n} w_{ij} z_j)$.

**Proof of Lemma A.1** Our proof follows Owen (1990, Pages 100–102). However, we obtain nonparametric (i.e. slower than $n^{1/2}$) rates of convergence for various terms as compared to Owen, who obtains parametric rates. Begin by observing that since the $\lambda_i$'s solve (2.6), we can write

$$\begin{aligned} 0 &= \sum_{j=1}^{n} \frac{w_{ij} z_j}{n + \lambda_i' z_j} = \sum_{j=1}^{n} \frac{(w_{ij} z_j / n)}{1 + (\lambda_i' z_j / n)} \\ &= \frac{1}{n} \sum_{j=1}^{n} w_{ij} z_j \{1 - \frac{\lambda_i' z_j}{n} + \frac{(\lambda_i' z_j / n)^2}{1 + (\lambda_i' z_j / n)}\} \\ &= \frac{1}{n} \sum_{j=1}^{n} w_{ij} z_j - \frac{1}{n^2} \hat{V}(x_i) \lambda_i + \frac{1}{n} \sum_{j=1}^{n} \frac{w_{ij} z_j (\lambda_i' z_j / n)^2}{1 + (\lambda_i' z_j / n)}. \end{aligned}$$

Therefore, since $\hat{V}(x_i)$ is invertible for large enough $n$, we have

$$\lambda_i = n \hat{V}^{-1}(x_i) \sum_{j=1}^{n} w_{ij} z_j + \hat{V}^{-1}(x_i) r_i^{(1)}, \quad r_i^{(1)} = n \sum_{j=1}^{n} \frac{w_{ij} z_j (\lambda_i' z_j / n)^2}{1 + (\lambda_i' z_j / n)}.$$

From Lemma B.1 we know that for $c_n \stackrel{\text{def}}{=} \sqrt{\frac{\log n}{n b_n^s}}$,

$$\max_{i \in I_*} \|r_i^{(1)}\| = n^{1+1/m} O_p(c_n^2).$$

In fact, because Assumption 3.3(*iii*) implies that $\|\hat{V}^{-1}(x_i)\|$ is uniformly bounded in probability on $S_*$, the above approximation for $\lambda_i$ holds uniformly in $i \in I_*$ and we can rewrite it as

$$\lambda_i = n \hat{V}^{-1}(x_i) \sum_{j=1}^{n} w_{ij} z_j + r_i^{(2)}, \quad \max_{i \in I_*} \|r_i^{(2)}\| = n^{1+1/m} O_p(c_n^2). \tag{A.1}$$

Now under our choice of $b_n$, the second result of Lemma B.1 ensures that $|\frac{\lambda_i' z_j}{n}| = o_p(1)$ uniformly in $i \in I_*$ and $j = 1, \ldots, n$. Therefore, for $i \in I_*$ and $1 \leq j \leq n$, the expansion

$$\log(1 + \frac{\lambda_i' z_j}{n}) = \frac{\lambda_i' z_j}{n} - \frac{1}{2} (\frac{\lambda_i' z_j}{n})^2 + \eta_{ij}$$

holds with the remainder term $\eta_{ij} = O_p(|\frac{\lambda_i' z_j}{n}|^3)$. Note that with probability approaching one as $n \uparrow \infty$, we can write $|\eta_{ij}| \leq B |\frac{\lambda_i' z_j}{n}|^3$ for some $B > 0$. $B$ does not depend upon $i$ and $j$ because $|\frac{\lambda_i' z_j}{n}|$ is asymptotically negligible in probability uniformly in $i \in I_*$ and $1 \leq j \leq n$. Using the above expansion and the expression for $\lambda_i$ in (A.1), some algebra shows that we can write

$$\mathrm{SELR} = \mathrm{T}_n - \frac{1}{n^2} \sum_{i=1}^{n} \mathbb{I}_i r_i^{(2)\prime} \hat{V}(x_i) r_i^{(2)} + 2 \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbb{I}_i w_{ij} \eta_{ij}.$$

But as we know that $\sup_{x_i \in S_*} \|\hat{V}(x_i)\| = O_p(1)$ and $\mathbb{I}_i \leq 1$, we can use the bounds for $r_i^{(2)}$ in (A.1) to see that

$$\frac{1}{n^2}\sum_{i=1}^{n} \mathbb{I}_i r_i^{(2)\prime} \hat{V}(x_i) r_i^{(2)} \leq n^{1+2/m} O_p(c_n^4) = O_p(\frac{\{\log n\}^2 n^{2/m}}{n b_n^{2s}}).$$

Also use the second result of Lemma B.1 to obtain

$$|\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{I}_i w_{ij}\eta_{ij}| \leq B \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} |\frac{\lambda_i' z_j}{n}|^3 \leq c\{n^{1/m} O_p(c_n)\}^3 \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}$$
$$= n^{3/m} O_p(c_n^3)\, n = O_p(\frac{\{\log n\}^{3/2} n^{\frac{3}{m}}}{\sqrt{n b_n^{3s}}}).$$

The conclusion follows by simplifying the terms inside the $O_p$ symbols. □

**Lemma A.2.** $n b_n^{2s}\, \mathrm{T}_{n,1} \overset{\mathrm{H}_0}{=} \mathcal{K}^2(0)\mathbb{E}[\mathbb{I}\{x_1 \in S_*\} z_1' \frac{V^{-1}(x_1)}{h^2(x_1)} z_1] + o_p(1)$.

**Proof of Lemma A.2**. Since $\hat{V}(x_i)$ converges element by element in probability to $V(x_i)$, $\hat{h}(x_i) \xrightarrow{p} h(x_i)$, and $h$ is bounded above zero on $S_*$,

$$\mathrm{T}_{n,1} = \mathcal{K}^2(0) \sum_{i=1}^{n} \mathbb{I}_i \frac{z_i' \hat{V}^{-1}(x_i) z_i}{\{\sum_{u=1}^{n} \mathcal{K}_{iu}\}^2} = \frac{\mathcal{K}^2(0)}{n^2 b_n^{2s}} \sum_{i=1}^{n} \mathbb{I}_i z_i' \hat{H}^{-1}(x_i) z_i$$
$$= \frac{\mathcal{K}^2(0)}{n^2 b_n^{2s}} \sum_{i=1}^{n} \mathbb{I}_i z_i' \{H^{-1}(x_i) + Rem_i\} z_i,$$

where $Rem_i$ represents a matrix of remainder terms which are asymptotically negligible in probability. Because each element of $Rem_i$ is $o_p(1)$ and the components of $z_i$ are bounded in probability,

$$\mathrm{T}_{n,1} = \frac{\mathcal{K}^2(0)}{n b_n^{2s}} \{\mathbb{E}[\mathbb{I}_1 z_1' H^{-1}(x_1) z_1] + o_p(1)\} + o_p(\frac{1}{n b_n^{2s}})$$

by an application of the WLLN. □

**Lemma A.3.** *Let* $\mathbb{E}\|z_1\|^m < \infty$ *for some* $m > 2$, *and choose* $b_n = n^{-\alpha}$ *for* $0 < \alpha < \frac{1}{s}(1 - \frac{2}{m})$. *Then under* $\mathrm{H}_0$,

$$\mathrm{T}_{n,2} = b_n^{-s}\{d\,\mathfrak{R}(\mathcal{K})\, vol(S_*) + O_p(\sqrt{\frac{\log n}{n b_n^s}} + b_n^2)\}.$$

**Proof of Lemma A.3**. Set $\delta = 0$ in the proof of Lemma A.9. □

**Lemma A.4.** *Let* $\mathbb{E}\|z_1\|^m < \infty$ *for some* $m > 2$ *and choose* $b_n = n^{-\alpha}$ *for* $0 < \alpha < \min\{\frac{1}{s}(1 - \frac{2}{m}), \frac{1}{2s}\}$. *Then under* $\mathrm{H}_0$, $\mathrm{T}_{n,3} = o_p(\frac{1}{\sqrt{n b_n^{2s}}})$.

**Proof of Lemma A.4**. Let $\hat{F}^{(lv)}(x_i)$ (resp. $G_n^{(lv)}(x_i)$) denote the $(lv)^{th}$ element of $\hat{H}^{-1}(x_i)$ (resp. $\frac{\{\mathbb{E}\hat{\Omega}(x_i)\}^{-1}}{\mathbb{E}\hat{h}(x_i)}$). From Lemma B.11 we know that

$$\sup_{x_i \in S_*} |\hat{F}^{(lv)}(x_i) - G_n^{(lv)}(x_i)| = O_p(c_n),$$

where $c_n = \sqrt{\frac{\log n}{nb_n^s}}$ and $c_n \to 0$ on choosing $b_n = n^{-\alpha}$ for $0 < \alpha < \frac{1}{s}(1 - \frac{2}{m})$. This result will be used later in the proof. Notice that since

$$\begin{aligned} \mathrm{T}_{n,3} &= \frac{\mathcal{K}(0)}{n^2 b_n^{2s}} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \mathbb{I}_i z_i' \hat{H}^{-1}(x_i) z_j \mathcal{K}_{ij} \\ &= \sum_{l=1}^{d}\sum_{v=1}^{d} \frac{\mathcal{K}(0)}{n^2 b_n^{2s}} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \mathbb{I}_i z_i^{(l)} \hat{F}^{(lv)}(x_i) z_j^{(v)} \mathcal{K}_{ij}, \end{aligned}$$

it suffices to show that

$$P_n \stackrel{\text{def}}{=} \frac{1}{n^{3/2} b_n^s} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \mathbb{I}_i z_i^{(l)} \hat{F}^{(lv)}(x_i) z_j^{(v)} \mathcal{K}_{ij} = o_p(1).$$

So write $P_n = P_n^{(1)} + P_n^{(2)}$ where

$$P_n^{(1)} = \frac{1}{n^{3/2} b_n^s} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \mathbb{I}_i z_i^{(l)} G_n^{(lv)}(x_i) z_j^{(v)} \mathcal{K}_{ij},$$

$$P_n^{(2)} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{I}_i z_i^{(l)} [\hat{F}^{(lv)}(x_i) - G_n^{(lv)}(x_i)] Q_{n,i}, \quad Q_{n,i} = \frac{1}{nb_n^s} \sum_{j=1, j\neq i}^{n} z_j^{(v)} \mathcal{K}_{ij}.$$

We look at $P_n^{(1)}$ and $P_n^{(2)}$ one by one. First note that by (D.2) and the Cauchy-Schwarz inequality

$$\mathbb{E}\{P_n^{(1)}\}^2 \le \frac{2}{n^3 b_n^{2s}} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \mathbb{E}\{z_i^{(l)} G_n^{(lv)}(x_i) z_j^{(v)} \mathcal{K}_{ij}\}^2,$$

where any remaining cross terms vanish since the random variables

$$\mathbb{I}_i z_i^{(l)} G_n^{(lv)}(x_i) z_j^{(v)} \mathcal{K}_{ij} \quad \text{and} \quad \mathbb{I}_i z_i^{(l)} G_n^{(lv)}(x_i) z_k^{(v)} \mathcal{K}_{ik}$$

are uncorrelated for $i \neq j \neq k$. Therefore, keeping in mind that $G_n^{(lv)}(x_i)$ is uniformly bounded in $x_i \in S_*$ for large enough $n$ (see the proof of Lemma B.11), it is straightforward to show that

$$\mathbb{E}\{P_n^{(1)}\}^2 \le \frac{c}{n^3 b_n^{2s}} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \mathbb{E}\{z_i^{(l)} z_j^{(v)} \mathcal{K}_{ij}\}^2 = O(\frac{1}{nb_n^s});$$

i.e. $P_n^{(1)} = O_p(\frac{1}{\sqrt{nb_n^s}})$. Next, by the Cauchy-Schwarz inequality we have

$$
\begin{aligned}
|P_n^{(2)}|^2 &\le \frac{1}{n}\sum_{i=1}^n [z_i^{(l)}]^2[\hat{F}^{(lv)}(x_i) - G_n^{(lv)}(x_i)]^2 \sum_{i=1}^n Q_{n,i}^2 \\
&\le \{\sup_{x_i\in S_*} |\hat{F}^{(lv)}(x_i) - G_n^{(lv)}(x_i)|\}^2 \,\frac{1}{n}\sum_{i=1}^n [z_i^{(l)}]^2 \sum_{i=1}^n Q_{n,i}^2 \\
&= O_p(c_n^2)\sum_{i=1}^n Q_{n,i}^2.
\end{aligned}
$$

But as $z_j^{(v)}\mathcal{K}_{ij}$ and $z_k^{(v)}\mathcal{K}_{ik}$ are uncorrelated for $i \ne j \ne k$, we can show

$$
\mathbb{E}Q_{n,i}^2 = \frac{1}{n^2 b_n^{2s}} \sum_{j=1,j\ne i}^n \mathbb{E}\{z_j^{(v)}\mathcal{K}_{ij}\}^2 = O(\frac{1}{nb_n^s}) \quad \text{(uniformly in } i \in \{1,\dots,n\}),
$$

which yields that $Q_{n,i}^2 = O_p(\frac{1}{nb_n^s})$ for all $i$. Therefore

$$
|P_n^{(2)}|^2 = O_p(c_n^2)\sum_{i=1}^n Q_{n,i}^2 = O_p(\frac{c_n^2}{b_n^s}),
$$

which implies $P_n^{(2)} = O_p(c_n b_n^{-s/2})$. Hence recalling that $c_n = \sqrt{\frac{\log n}{nb_n^s}}$,

$$
\begin{aligned}
P_n &= O_p(\frac{1}{\sqrt{nb_n^s}}) + O_p(c_n b_n^{-s/2}) = O_p(\frac{1}{\sqrt{nb_n^s}}) + O_p(\sqrt{\frac{\log n}{nb_n^{2s}}}) \\
&= O_p(\sqrt{\frac{\log n}{nb_n^{2s}}}) = o_p(1).
\end{aligned}
$$

The last equality follows if $b_n = n^{-\alpha}$ for $0 < \alpha < \min\{\frac{1}{s}(1-\frac{2}{m}), \frac{1}{2s}\}$. □

**Lemma A.5.** *Let* $\mathbb{E}\|z_1\|^m < \infty$ *for some* $m > 2$ *and choose* $b_n = n^{-\alpha}$ *for* $0 < \alpha < \min\{\frac{1}{s}(1-\frac{2}{m}), \frac{1}{2s}\}$. *Then under* $\mathrm{H}_0$, $\mathrm{T}_{n,4} = o_p(\frac{1}{\sqrt{nb_n^{2s}}})$.

**Proof of Lemma A.5**. Same as the proof of Lemma A.4. □

**Lemma A.6.** *Let* $\mathbb{E}\|z_1\|^m < \infty$ *for some* $m > 4$ *and choose* $b_n = n^{-\alpha}$ *for* $0 < \alpha < \frac{1}{2s}(1-\frac{4}{m})$. *Then under* $\mathrm{H}_0$, $b_n^{s/2}\mathrm{T}_{n,5} \xrightarrow{d} \mathrm{N}(0, 2d\mathcal{K}^{**}vol(S_*))$.

**Proof of Lemma A.6** Write $\mathrm{T}_{n,5} = \mathrm{T}_{n,5}^* + (\mathrm{T}_{n,5} - \mathrm{T}_{n,5}^*)$ where

$$
\text{(A.2)} \quad \mathrm{T}_{n,5}^* = \frac{\tilde{\mathrm{T}}_{n,5}^*}{n^2 b_n^{2s}}, \qquad \tilde{\mathrm{T}}_{n,5}^* = \sum_{i=1}^n \sum_{j=1,j\ne i}^n \sum_{t=1,t\ne j\ne i}^n \mathbb{I}_i \mathcal{K}_{ij} z_j' H^{-1}(x_i) z_t \mathcal{K}_{it}.
$$

As $b_n^{s/2}\{\mathrm{T}_{n,5} - \mathrm{T}_{n,5}^*\} = o_p(1)$ from Lemma B.2 provided that $b_n = n^{-\alpha}$ for $0 < \alpha < \frac{1}{s}(1-\frac{2}{m})$, it suffices to show that under $\mathrm{H}_0$

$$
b_n^{s/2}\mathrm{T}_{n,5}^* \xrightarrow{d} \mathrm{N}(0, 2d\mathcal{K}^{**}vol(S_*)).
$$

To do so we will use a CLT for generalized quadratic forms due to de Jong (1987). First notice that since the only restriction on the summation signs in (A.2) is that $t \neq j \neq i$, we can change the order of summation in (A.2) to write

$$\begin{aligned}\tilde{\mathrm{T}}^*_{n,5} &= \sum_{t=1}^{n} \sum_{j=1, j\neq t}^{n} \sum_{i=1, i\neq j\neq t}^{n} \mathbb{I}_i \mathcal{K}_{ij} z_j' H^{-1}(x_i) z_t \mathcal{K}_{it} \\ &= \sum_{t=1}^{n} \sum_{j=1, j\neq t}^{n} z_t' A_{tjn} z_j, \qquad A_{tjn} = \sum_{i=1, i\neq j\neq t}^{n} \mathbb{I}_i \mathcal{K}_{ij} H^{-1}(x_i) \mathcal{K}_{it}.\end{aligned}$$

Next define $W_{tjn} = z_t' A_{tjn} z_j + z_j' A_{tjn} z_t = 2 z_t' A_{tjn} z_j$, and use Lemma B.3 to verify that $W_{tjn}$ is clean[7]. Using de Jong's notation we can then write $\tilde{\mathrm{T}}^*_{n,5} = \sum_{t=1}^{n-1} \sum_{j=t+1}^{n} W_{tjn}$. Let us now find $s_n^2$, the variance of $\tilde{\mathrm{T}}^*_{n,5}$. So using (D.3) we can write

$$s_n^2 = var \tilde{\mathrm{T}}^*_{n,5} = \sum_{t=1}^{n-1} \sum_{j=t+1}^{n} \mathbb{E} W_{tjn}^2 = 4 \sum_{t=1}^{n-1} \sum_{j=t+1}^{n} \mathbb{E}\{z_t' A_{tjn} z_j\}^2,$$

where any cross terms vanish due to the orthogonality of $W_{tjn}$ and $W_{tkn}$ for $t \neq j \neq k$ (see Remark B.1). Then using Lemma B.4 it follows that

$$s_n^2 = n(n-1)(n-2)(n-3) 2 d b_n^{3s} \mathcal{K}^{**} vol(S_*)\{1 + o(1)\}.$$

Next, as in de Jong (1987, Page 266), define the following terms:

$$G_I = \sum_{t=1}^{n-1} \sum_{j=t+1}^{n} \mathbb{E} W_{tjn}^4,$$

$$G_{II} = \sum_{t=1}^{n-2} \sum_{j=t+1}^{n-1} \sum_{k=j+1}^{n} (\mathbb{E} W_{tjn}^2 W_{tkn}^2 + \mathbb{E} W_{jtn}^2 W_{jkn}^2 + \mathbb{E} W_{ktn}^2 W_{kjn}^2),$$

$$\begin{aligned}G_{IV} = \sum_{t=1}^{n-3} \sum_{j=t+1}^{n-2} \sum_{k=j+1}^{n-1} \sum_{l=k+1}^{n} (&\mathbb{E} W_{tjn} W_{tkn} W_{ljn} W_{lkn} + \mathbb{E} W_{tjn} W_{tln} W_{kjn} W_{kln} \\ &+ \mathbb{E} W_{tkn} W_{tln} W_{jkn} W_{jln}).\end{aligned}$$

From Lemmas B.5, B.6, and B.7, we can see that:

$$\frac{G_I}{s_n^4} = \frac{O(n^{4+\frac{8}{m}} b_n^{2s})}{n^8 b_n^{6s} O(1)} = O\left(\frac{1}{n^{4-\frac{8}{m}} b_n^{4s}}\right), \quad \frac{G_{II}}{s_n^4} = \frac{O(n^{5+\frac{8}{m}} b_n^{2s})}{n^8 b_n^{6s} O(1)} = O\left(\frac{1}{n^{3-\frac{8}{m}} b_n^{4s}}\right),$$

$$\frac{G_{IV}}{s_n^4} = \frac{O(n^{6+\frac{8}{m}} b_n^{2s})}{n^8 b_n^{6s} O(1)} = O\left(\frac{1}{n^{2-\frac{8}{m}} b_n^{4s}}\right).$$

[7] According to de Jong (1987, Page 263), $W_{tjn}$ is said to be "clean" if $\mathbb{E}(W_{tjn}|x_t, z_t) = \mathbb{E}(W_{tjn}|x_j, z_j) = 0$ a.s. for all $1 \leq t, j \leq n$.

If we let $b_n = n^{-\alpha}$, where $0 < \alpha < \frac{1}{2s}(1 - \frac{4}{m})$, we get $G_I, G_{II}, G_{IV} = o(s_n^4)$. Hence, from de Jong (1987, Proposition 3.2, Page 267), we have $s_n^{-1}\tilde{\mathrm{T}}^*_{n,5} \xrightarrow[\mathbb{E}(z|x)=0]{d} \mathrm{N}(0,1)$. So using Slutsky's lemma we obtain

$$b_n^{s/2}\mathrm{T}^*_{n,5} = \frac{\tilde{\mathrm{T}}^*_{n,5}}{n^2 b_n^{3s/2}} \xrightarrow[\mathbb{E}(z|x)=0]{d} \mathrm{N}(0, 2d\mathcal{K}^{**}vol(S_*)). \quad \square$$

**Lemma A.7.** *Let* $\mathbb{E}\|z_1\|^m < \infty$ *for some* $m > 2$ *and choose* $b_n = n^{-\alpha}$ *for* $0 < \alpha < \frac{1}{s}(1-\frac{2}{m})$. *Then* $\mathrm{SELR} \overset{\mathrm{H}_{1n}}{=} \mathrm{T}_n + O_p(\{\frac{\log n}{n^{\frac{1}{2}-\frac{1}{m}} b_n^s}\}^2) + O_p(\{\frac{\log n}{n^{\frac{1}{3}-\frac{2}{m}} b_n^s}\}^{3/2})$.

**Proof of Lemma A.7**. Following the details in the first half of the proof of Lemma A.1, we can write

$$\lambda_i = n\hat{V}^{-1}(x_i)\sum_{j=1}^{n} w_{ij}z_j + \hat{V}^{-1}(x_i)r_i^{(1)}(\mathrm{H}_{1n}),$$

where $r_i^{(1)}(\mathrm{H}_{1n}) = n\sum_{j=1}^n \frac{w_{ij}z_j(\lambda_i' z_j/n)^2}{1+(\lambda_i' z_j/n)}$. Getting this far only required algebraic manipulations; i.e. we did not use the fact that $\mathbb{E}(z|x) \overset{\mathrm{H}_{1n}}{=} \epsilon_n\delta(x)$, where $\epsilon_n \overset{\mathrm{def}}{=} n^{-1/2}b_n^{-s/4}$ for convenience. We do so now. Using the algebra that led to (B.2), we can see that

$$\|r_i^{(1)}(\mathrm{H}_{1n})\| \le n^{1/m}\rho_i \sum_{j=1}^{n} w_{ij}\xi_i' z_j,$$

where we have followed the notation of Lemma B.1 to write $\lambda_i = \rho_i\xi$ for $\rho_i > 0$ and $\xi_i \in \mathbb{S}$. Now let $c_n \overset{\mathrm{def}}{=} \sqrt{\frac{\log n}{nb_n^s}}$, and observe that $c_n \to 0$ on letting $b_n = n^{-\alpha}$ for $0 < \alpha < \frac{1}{s}(1 - \frac{2}{m})$. From Newey (1994, Lemma B.1, Page 250) we know that

$$\sup_{x_i\in S_*} |\frac{1}{nb_n^s}\sum_{j=1}^{n}\mathcal{K}_{ij}z_j - \mathbb{E}\{\frac{1}{nb_n^s}\sum_{j=1}^{n}\mathcal{K}_{ij}z_j\}| \overset{\text{elt. by elt.}}{=} O_p(c_n),$$

$$\sup_{x_i\in S_*} |\hat{h}(x_i) - \mathbb{E}\hat{h}(x_i)| = O_p(c_n),$$

$$\sup_{x_i\in S_*} |\mathbb{E}\hat{h}(x_i) - h(x_i)| = O_p(b_n^2).$$

As $h$ is bounded away from zero on $S_*$, the last result implies that $\mathbb{E}\hat{h}(x_i)$ is also bounded away from zero on $S_*$ for large enough $n$. Therefore, using Lemma C.1 and the first two results,

$$\sup_{x_i\in S_*} |\sum_{j=1}^{n} w_{ij}z_j - \frac{\mathbb{E}\{\frac{1}{nb_n^s}\sum_{j=1}^n \mathcal{K}_{ij}z_j\}}{\mathbb{E}\hat{h}(x_i)}| \overset{\text{elt. by elt.}}{=} O_p(c_n).$$

But following Assumption 5.1, it is straightforward to see that under $\mathrm{H}_{1n}$

$$\begin{aligned}\mathbb{E}\{\frac{1}{nb_n^s}\sum_{j=1}^n \mathcal{K}_{ij}z_j\} &= \mathbb{E}\{\frac{\epsilon_n}{nb_n^s}\sum_{j=1}^n \mathcal{K}_{ij}\delta(x_j)\} \\ &= \epsilon_n \int_{S_\mathcal{K}} \mathcal{K}(u)\delta(x_i - b_n u)h(x_i - b_n u)\,du \\ &\overset{\text{elt. by elt.}}{=} O(\epsilon_n) \qquad \text{(uniformly in } x_i \in S_*),\end{aligned}$$

which implies that

$$\text{(A.3)}\quad \sum_{j=1}^n w_{ij}\xi' z_j \overset{\mathrm{H}_{1n}}{=} O_p(\max\{\epsilon_n, c_n\}) = O_p(c_n) \quad \text{(uniformly in } x_i \in S_*).$$

Hence $\|r_i^{(1)}(\mathrm{H}_{1n})\| \le n^{1/m}\rho_i O_p(c_n)$, which is identical to the result in (B.3) obtained under $\mathrm{H}_0$. Furthermore, because $\|\hat{V}^{-1}(x_i)\|$ is uniformly bounded in probability on $S_*$, we can use the approach described in the latter half of Lemma B.1 to show that $\max_{i\in I_*}\, \rho_i = nO_p(c_n)$. Thus we obtain

$$\lambda_i \overset{\mathrm{H}_{1n}}{=} n\hat{V}^{-1}(x_i)\sum_{j=1}^n w_{ij}z_j + r_i^{(2)}(\mathrm{H}_{1n}), \quad \max_{i\in I_*}\|\, r_i^{(2)}(\mathrm{H}_{1n})\| = n^{1+1/m}O_p(c_n^2).$$

This approximation for $\lambda_i$ is identical to the one obtained in (A.1) under the null hypothesis. In a similar manner we can use (A.3) to show that all remaining approximations in the proof of Lemma A.1 stay unchanged even under $\mathrm{H}_{1n}$. The desired result follows. □

**Lemma A.8.** $nb_n^{2s}\,\mathrm{T}_{n,1} \overset{\mathrm{H}_{1n}}{=} \mathcal{K}^2(0)\mathbb{E}[\mathbb{I}\{x_1 \in S_*\}z_1' \frac{V^{-1}(x_1)}{h^2(x_1)} z_1] + o_p(1)$.

**Proof of Lemma A.8**. Same as Lemma A.2. □

**Lemma A.9.** *Let* $\mathbb{E}\|\tilde{z}_1\|^m < \infty$ *for some* $m > 2$ *and choose* $b_n = n^{-\alpha}$ *for* $0 < \alpha < \frac{1}{s}(1 - \frac{2}{m})$*. Then under* $\mathrm{H}_{1n}$,

$$\mathrm{T}_{n,2} = b_n^{-s}\{d\,\mathfrak{R}(\mathcal{K})\,vol(S_*) + O_p(\sqrt{\frac{\log n}{nb_n^s}} + b_n^2)\}.$$

**Proof of Lemma A.9**. For notational convenience define $\epsilon_n = n^{-1/2} b_n^{-s/4}$. Now write $\mathrm{T}_{n,2} = (A_1) + (A_2) + (A_3) + (A_4)$, where

$$(A_1) = \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \mathbb{I}_i w_{ij}^2 \tilde{z}_j' \hat{V}^{-1}(x_i) \tilde{z}_j$$

$$(A_2) = \epsilon_n \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \mathbb{I}_i w_{ij}^2 \tilde{z}_j' \hat{V}^{-1}(x_i) \delta(x_j)$$

$$(A_3) = \epsilon_n \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \mathbb{I}_i w_{ij}^2 \delta'(x_j) \hat{V}^{-1}(x_i) \tilde{z}_j$$

$$(A_4) = \epsilon_n^2 \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \mathbb{I}_i w_{ij}^2 \delta'(x_j) \hat{V}^{-1}(x_i) \delta(x_j).$$

Let us first look at $(A_1)$. So begin by observing that

$$(A_1) = \frac{1}{n^2 b_n^{2s}} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \mathbb{I}_i \mathcal{K}_{ij}^2 \tilde{z}_j' \hat{H}^{-1}(x_i) \tilde{z}_j = (A_1)_a + (A_1)_b, \quad \text{where}$$

$$(A_1)_a = \frac{1}{n^2 b_n^{2s}} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \mathbb{I}_i \mathcal{K}_{ij}^2 \tilde{z}_j' H^{-1}(x_i) \tilde{z}_j$$

$$(A_1)_b = \frac{1}{n^2 b_n^{2s}} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \mathbb{I}_i \mathcal{K}_{ij}^2 \tilde{z}_j' \{\hat{H}^{-1}(x_i) - H^{-1}(x_i)\} \tilde{z}_j.$$

Next note that

$$|\mathcal{K}_{ij}^2 \tilde{z}_j' \{\hat{H}^{-1}(x_i) - H^{-1}(x_i)\} \tilde{z}_j| \leq \mathcal{K}_{ij}^2 \|\tilde{z}_j\|^2 \, \|\hat{H}^{-1}(x_i) - H^{-1}(x_i)\|.$$

But letting $\tau_n \stackrel{\text{def}}{=} \sqrt{\frac{\log n}{n b_n^s}} + b_n^2$, from Lemma B.10 we know that

$$\sup_{x_i \in S_*} |\hat{H}^{-1}(x_i) - H^{-1}(x_i)| \stackrel{\text{elt. by elt.}}{=} O_p(\tau_n),$$

and $\tau_n \to 0$ on choosing $b_n = n^{-\alpha}$ for $0 < \alpha < \frac{1}{s}(1 - \frac{2}{m})$. Therefore, since

$$\begin{aligned}
(A_1)_b &\leq \frac{1}{n^2 b_n^{2s}} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \mathcal{K}_{ij}^2 \|\tilde{z}_j\|^2 \|\hat{H}^{-1}(x_i) - H^{-1}(x_i)\| \\
&\leq \frac{O_p(\tau_n)}{n b_n^s} \sum_{i=1}^{n} \{ \frac{1}{n b_n^s} \sum_{j=1, j\neq i}^{n} \mathcal{K}_{ij}^2 \|\tilde{z}_j\|^2 \} \\
&= \frac{O_p(\tau_n)}{b_n^s} \, O_p(1) = O_p(\frac{\tau_n}{b_n^s}),
\end{aligned}$$

we get $(A_1) = (A_1)_a + O_p(\frac{\tau_n}{b_n^s})$. Now write

$$\begin{aligned}
(A_1)_a &= \frac{1}{nb_n^s}\, tr \sum_{i=1}^{n} \mathbb{I}_i \{ \frac{1}{nb_n^s} \sum_{j=1, j\neq i}^{n} \mathcal{K}_{ij}^2 \tilde{z}_j \tilde{z}_j' \}\, H^{-1}(x_i) \\
&= \frac{1}{nb_n^s}\, tr \sum_{i=1}^{n} \mathbb{I}_i \{ \mathfrak{R}(\mathcal{K}) \mathbb{E}(\tilde{z}\tilde{z}'|x_i) h(x_i) + R^{(A)}(x_i) \}\, H^{-1}(x_i),
\end{aligned}$$

where $R^{(A)}(x_i) \overset{\text{elt. by elt.}}{=} O_p(b_n^2)$ follows from the consistency of kernel estimators. Because $\mathbb{E}(\tilde{z}_i \tilde{z}_i' | x_i) = V(x_i)$ and $V^{-1}(x_i)$ is element by element uniformly bounded on $S_*$, the previous equation reduces to

$$\begin{aligned}
(A_1)_a &= \frac{1}{nb_n^s}\, tr \sum_{i=1}^{n} \{ \frac{\mathfrak{R}(\mathcal{K}) \mathbb{I}_i}{h(x_i)} I_{d\times d} + \mathbb{I}_i R^{(A)}(x_i) H^{-1}(x_i) \} \\
&= \frac{d\,\mathfrak{R}(\mathcal{K})}{nb_n^s} \sum_{i=1}^{n} \frac{\mathbb{I}_i}{h(x_i)} + \frac{1}{nb_n^s} \sum_{i=1}^{n} \mathbb{I}_i\, tr\, R^{(A)}(x_i) H^{-1}(x_i) \\
&= \frac{d\,\mathfrak{R}(\mathcal{K})}{nb_n^s} \sum_{i=1}^{n} \frac{\mathbb{I}_i}{h(x_i)} + \frac{O_p(b_n^2)}{b_n^s}.
\end{aligned}$$

By the CLT we know that

$$n^{1/2} \left( \frac{\frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_i h^{-1}(x_i) - \mathbb{E}\{\mathbb{I}_1 h^{-1}(x_1)\}}{\sqrt{var\{\mathbb{I}_1 h^{-1}(x_1)\}}} \right) = O_p(1);$$

i.e.

$$\frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_i h^{-1}(x_i) = vol(S_*) + O_p(n^{-1/2}).$$

Using this approximation we have

$$(A_1)_a = \frac{1}{b_n^s} \{ d\mathfrak{R}(\mathcal{K}) vol(S_*) + O_p(n^{-1/2}) + O_p(b_n^2) \}.$$

Combining the results for $(A_1)_a$ and $(A_1)_b$ we get that

$$\begin{aligned}
(A_1) &= \frac{1}{b_n^s} \{ d\mathfrak{R}(\mathcal{K}) vol(S_*) + O_p(n^{-1/2}) + O_p(b_n^2) \} + O_p(\frac{\tau_n}{b_n^s}) \\
&= \frac{1}{b_n^s} \{ d\mathfrak{R}(\mathcal{K}) vol(S_*) + O_p(n^{-1/2}) + O_p(b_n^2) + O_p(\tau_n) \} \\
&= \frac{1}{b_n^s} \{ d\mathfrak{R}(\mathcal{K}) vol(S_*) + O_p(\tau_n) \}.
\end{aligned}$$

Now let us look at $(A_2)$. Observe that we can write

$$(A_2) = \frac{\epsilon_n}{nb_n^s}\, tr \sum_{i=1}^{n} \mathbb{I}_i \{ \frac{1}{nb_n^s} \sum_{j=1, j\neq i}^{n} \mathcal{K}_{ij}^2 \delta(x_j) \tilde{z}_j' \}\, \hat{H}^{-1}(x_i).$$

But since $(i)$ $\frac{1}{nb_n^s}\sum_{j=1,j\neq i}^n \mathcal{K}_{ij}^2\delta(x_j)\tilde{z}_j' \xrightarrow{p} O_{d\times d}$ because $\mathbb{E}(\tilde{z}_j|x_j) = 0$, and $(ii)$ $\hat{H}^{-1}(x_i)$ is element by element $O_p(1)$ on $S_*$, we get

$$(A_2) = \frac{\epsilon_n}{nb_n^s}\, o_p(n) = o_p(\frac{\epsilon_n}{b_n^s}).$$

Similarly, we can also show that $(A_3) = o_p(\frac{\epsilon_n}{b_n^s})$. Finally, observe that

$$(A_4) = \frac{\epsilon_n^2}{nb_n^s}\, tr\sum_{i=1}^n \mathbb{I}_i\{\frac{1}{nb_n^s}\sum_{j=1,j\neq i}^n \mathcal{K}_{ij}^2\delta(x_j)\delta'(x_j)\}\,\hat{H}^{-1}(x_i).$$

But $\hat{H}^{-1}(x_i)$ is element by element $O_p(1)$ on $S_*$, and from Assumption 5.1 we know that $\frac{1}{nb_n^s}\sum_{j=1,j\neq i}^n \mathcal{K}_{ij}^2\delta(x_j)\delta'(x_j)$ is also element by element $O_p(1)$ on $S_*$. This yields

$$(A_4) = \frac{\epsilon_n^2}{nb_n^s}\, O_p(n) = O_p(\frac{\epsilon_n^2}{b_n^s}).$$

Combining the results for $(A_1)$—$(A_4)$ we get that under $\mathrm{H}_{1n}$,

$$\mathrm{T}_{n,2} = \frac{1}{b_n^s}\{d\mathfrak{R}(\mathcal{K})vol(S_*) + O_p(\tau_n)\} + o_p(\frac{\epsilon_n}{b_n^s}) + o_p(\frac{\epsilon_n}{b_n^s}) + O_p(\frac{\epsilon_n^2}{b_n^s});$$

i.e. $\mathrm{T}_{n,2} = b_n^{-s}\{d\,\mathfrak{R}(\mathcal{K})\,vol(S_*) + O_p(\tau_n)\}$. □

**Lemma A.10.** *Let $\mathbb{E}\|\tilde{z}_1\|^m < \infty$ for some $m > 2$ and choose $b_n = n^{-\alpha}$ for $0 < \alpha < \min\{\frac{1}{s}(1-\frac{2}{m}), \frac{1}{2s}\}$. Then under $\mathrm{H}_{1n}$, $\mathrm{T}_{n,3} = O_p(\frac{1}{\sqrt{nb_n^{5s/2}}})$.*

**Proof of Lemma A.10**. Define $\epsilon_n = n^{-1/2}b_n^{-s/4}$. Under $\mathrm{H}_{1n}$ we can write $\mathrm{T}_{n,3} = (B_1) + (B_2) + (B_3) + (B_4)$, where

$$(B_1) = \frac{\mathcal{K}(0)}{n^2b_n^{2s}}\sum_{i=1}^n\sum_{j=1,j\neq i}^n \mathbb{I}_i\mathcal{K}_{ij}\tilde{z}_i'\hat{H}^{-1}(x_i)\tilde{z}_j$$

$$(B_2) = \frac{\epsilon_n\mathcal{K}(0)}{n^2b_n^{2s}}\sum_{i=1}^n\sum_{j=1,j\neq i}^n \mathbb{I}_i\mathcal{K}_{ij}\tilde{z}_i'\hat{H}^{-1}(x_i)\delta(x_j)$$

$$(B_3) = \frac{\epsilon_n\mathcal{K}(0)}{n^2b_n^{2s}}\sum_{i=1}^n\sum_{j=1,j\neq i}^n \mathbb{I}_i\mathcal{K}_{ij}\delta'(x_i)\hat{H}^{-1}(x_i)\tilde{z}_j$$

$$(B_4) = \frac{\epsilon_n^2\mathcal{K}(0)}{n^2b_n^{2s}}\sum_{i=1}^n\sum_{j=1,j\neq i}^n \mathbb{I}_i\mathcal{K}_{ij}\delta'(x_i)\hat{H}^{-1}(x_i)\delta(x_j).$$

In Lemma A.4 we had shown that under $\mathrm{H}_0$

$$\mathcal{K}(0)\sum_{i=1}^n\sum_{j=1,j\neq i}^n \mathbb{I}_i z_i'\hat{H}^{-1}(x_i)z_j\mathcal{K}_{ij} = o_p(n^{3/2}b_n^s),$$

provided we chose $b_n = n^{-\alpha}$ for $0 < \alpha < \min\{\frac{1}{s}(1-\frac{2}{m}), \frac{1}{2s}\}$. The only properties we had used to obtain this result were that $\mathbb{E}(z_i|x_i) = 0$ and that the $z_i$'s were iid. Since the same properties hold for $\tilde{z}_i$ under $\mathrm{H}_{1n}$, we also have $(B_1) \overset{\mathrm{H}_{1n}}{=} o_p(\frac{1}{\sqrt{nb_n^{2s}}})$ if the bandwidth is chosen as described. Next write

$$(B_2) = \frac{\epsilon_n \mathcal{K}(0)}{nb_n^s} \sum_{i=1}^{n} \mathbb{I}_i \tilde{z}_i' \hat{H}^{-1}(x_i) \{ \frac{1}{nb_n^s} \sum_{j=1, j\neq i}^{n} \mathcal{K}_{ij} \delta(x_j) \} = \frac{\epsilon_n}{nb_n^s} O_p(n),$$

where the second equality follows because: $(i)$ $\frac{1}{nb_n^s} \sum_{j=1,j\neq i}^{n} \mathcal{K}_{ij}\delta(x_j) \overset{\text{elt. by elt.}}{=} O_p(1)$, $(ii)$ $\mathbb{E}\|\tilde{z}\|^2 < \infty$, and $(iii)$ $\hat{H}^{-1}(x_i)$ is element by element uniformly bounded in probability on $S_*$. Since $\epsilon_n \overset{\text{def}}{=} n^{-1/2} b_n^{-s/4}$, we obtain that $(B_2) \overset{\mathrm{H}_{1n}}{=} O_p(\frac{1}{\sqrt{nb_n^{5s/2}}})$. Similarly, we can show that $(B_3) \overset{\mathrm{H}_{1n}}{=} O_p(\frac{1}{\sqrt{nb_n^{5s/2}}})$. Finally, we have

$$(B_4) = \frac{\epsilon_n^2 \mathcal{K}(0)}{nb_n^s} \sum_{i=1}^{n} \mathbb{I}_i \delta'(x_i) \hat{H}^{-1}(x_i) \{ \frac{1}{nb_n^s} \sum_{j=1, j\neq i}^{n} \mathcal{K}_{ij} \delta(x_j) \} = \frac{\epsilon_n^2}{nb_n^s} O_p(n),$$

where the second equality follows from $(i)$ and $(iii)$ as listed above and the fact that $\delta(x_i)$ is element by element $O_p(1)$. Therefore, $(B_4) \overset{\mathrm{H}_{1n}}{=} O_p(\frac{1}{nb_n^{3s/2}})$ and the lemma stands proved upon combining the results for $(B_1)$–$(B_4)$ and using the chosen bandwidth. □

**Lemma A.11.** *Let* $\mathbb{E}\|\tilde{z}_1\|^m < \infty$ *for some* $m > 2$ *and choose* $b_n = n^{-\alpha}$ *for* $0 < \alpha < \min\{\frac{1}{s}(1 - \frac{2}{m}), \frac{1}{2s}\}$. *Then under* $\mathrm{H}_{1n}$, $\mathrm{T}_{n,4} = O_p(\frac{1}{\sqrt{nb_n^{5s/2}}})$.

**Proof of Lemma A.11**. Same as the proof of Lemma A.10. □

**Lemma A.12.** *Let* $\mathbb{E}\|\tilde{z}_1\|^m < \infty$ *for some* $m > 4$ *and choose* $b_n = n^{-\alpha}$ *for* $0 < \alpha < \frac{1}{2s}(1 - \frac{4}{m})$. *Also let* $\mu = \mathbb{E}[\mathbb{I}\{x_1 \in S_*\}\delta'(x_1)V^{-1}(x_1)\delta(x_1)]$ *and* $\sigma^2 = 2d\mathcal{K}^{**}vol(S_*)$. *Then under* $\mathrm{H}_{1n}$, $b_n^{s/2}\mathrm{T}_{n,5} \xrightarrow{d} \mathrm{N}(\mu, \sigma^2)$.

**Proof of Lemma A.12**. Define $\epsilon_n = n^{-1/2} b_n^{-s/4}$. Under $\mathrm{H}_{1n}$ we can write $\mathrm{T}_{n,5} = (C_1) + (C_2) + (C_3) + (C_4)$, where

$$(C_1) = \frac{1}{n^2 b_n^{2s}} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \sum_{t=1, t\neq j\neq i}^{n} \mathbb{I}_i \mathcal{K}_{ij} \tilde{z}_j' \hat{H}^{-1}(x_i) \tilde{z}_t \mathcal{K}_{it}$$

$$(C_2) = \frac{\epsilon_n}{n^2 b_n^{2s}} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \sum_{t=1, t\neq j\neq i}^{n} \mathbb{I}_i \mathcal{K}_{ij} \tilde{z}_j' \hat{H}^{-1}(x_i) \delta(x_t) \mathcal{K}_{it}$$

$$(C_3) = \frac{\epsilon_n}{n^2 b_n^{2s}} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \sum_{t=1, t\neq j\neq i}^{n} \mathbb{I}_i \mathcal{K}_{ij} \delta'(x_j) \hat{H}^{-1}(x_i) \tilde{z}_t \mathcal{K}_{it}$$

$$(C_4) = \frac{\epsilon_n^2}{n^2 b_n^{2s}} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \sum_{t=1, t\neq j\neq i}^{n} \mathbb{I}_i \mathcal{K}_{ij} \delta'(x_j) \hat{H}^{-1}(x_i) \delta(x_t) \mathcal{K}_{it}.$$

Note that $(C_1)$ would be identical to $\mathrm{T}^*_{n,5}$ (which is defined in Lemma A.6) were it not for the fact that the former is a function of $\tilde{z}$ instead of $z$. Therefore, we can once again use Lemma A.6 to show that $b_n^{s/2}(C_1) \xrightarrow{d} \mathrm{N}(0,\sigma^2)$ provided $b_n = n^{-\alpha}$ for $0 < \alpha < \frac{1}{2s}(1-\frac{4}{m})$. Next write $(C_2) = (C_2)_a + (C_2)_b$, where

$$(C_2)_a = \frac{\epsilon_n}{n^2 b_n^{2s}} \sum_{i=1}^n \sum_{j=1,j\neq i}^n \sum_{t=1,t\neq j\neq i}^n \mathbb{I}_i \mathcal{K}_{ij} \tilde{z}_j' H^{-1}(x_i)\delta(x_t)\mathcal{K}_{it},$$

$$(C_2)_b = \frac{\epsilon_n}{n^2 b_n^{2s}} \sum_{i=1}^n \sum_{j=1,j\neq i}^n \sum_{t=1,t\neq j\neq i}^n \mathbb{I}_i \mathcal{K}_{ij} \tilde{z}_j' \{\hat{H}^{-1}(x_i) - H^{-1}(x_i)\}\delta(x_t)\mathcal{K}_{it}.$$

From Lemmas B.8 and B.9 it follows that both $b_n^{s/2}(C_2)_a$ and $b_n^{s/2}(C_2)_b$ are asymptotically negligible in probability. Therefore, $b_n^{s/2}(C_2) \stackrel{\mathrm{H}_{1n}}{=} o_p(1)$. Similarly, we can show that $b_n^{s/2}(C_3) \stackrel{\mathrm{H}_{1n}}{=} o_p(1)$. To analyze $(C_4)$ recall that $\epsilon_n \stackrel{\text{def}}{=} n^{-1/2} b_n^{-s/4}$ and write

$$\begin{aligned} b_n^{s/2}(C_4) &= \frac{1}{n^2 b_n^s} \sum_{i=1}^n \sum_{j=1,j\neq i}^n \mathbb{I}_i \mathcal{K}_{ij} \delta'(x_j) \hat{H}^{-1}(x_i) \{ \frac{1}{n b_n^s} \sum_{t=1,t\neq j\neq i}^n \delta(x_t)\mathcal{K}_{it} \} \\ &= \frac{1}{n^2 b_n^s} \sum_{i=1}^n \sum_{j=1,j\neq i}^n \mathbb{I}_i \mathcal{K}_{ij} \delta'(x_j) \hat{H}^{-1}(x_i) \{ \delta(x_i) h(x_i) + r_i^{(\alpha)} \}, \end{aligned}$$

where $r_i^{(\alpha)} \stackrel{\text{elt. by elt.}}{=} o_p(1)$, and the second equality follows from the consistency of kernel estimators. Hence let $b_n^{s/2}(C_4) = (C_4)_a + (C_4)_b$, where

$$(C_4)_a = \frac{1}{n^2 b_n^s} \sum_{i=1}^n \sum_{j=1,j\neq i}^n \mathbb{I}_i \mathcal{K}_{ij} \delta'(x_j) \hat{H}^{-1}(x_i) \delta(x_i) h(x_i)$$

$$(C_4)_b = \frac{1}{n^2 b_n^s} \sum_{i=1}^n \sum_{j=1,j\neq i}^n \mathbb{I}_i \mathcal{K}_{ij} \delta'(x_j) \hat{H}^{-1}(x_i) r_i^{(\alpha)}.$$

Once again, use the consistency of kernel estimators to see that

$$\begin{aligned} (C_4)_a &= \frac{1}{n} \sum_{i=1}^n \mathbb{I}_i \{ \frac{1}{n b_n^s} \sum_{j=1,j\neq i}^n \mathcal{K}_{ij} \delta'(x_j) \} \hat{H}^{-1}(x_i) \delta(x_i) h(x_i) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{I}_i \{ \delta'(x_i) h(x_i) + r_i^{(\beta)'} \} \hat{H}^{-1}(x_i) \delta(x_i) h(x_i), \end{aligned}$$

where $r_i^{(\beta)} \stackrel{\text{elt. by elt.}}{=} o_p(1)$. Similarly, letting $r_i^{(\gamma)} \stackrel{\text{def}}{=} \frac{1}{n b_n^s} \sum_{j=1,j\neq i}^n \mathcal{K}_{ij} \delta'(x_j)$ we have

$$(C_4)_b = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_i r_i^{(\gamma)'} \hat{H}^{-1}(x_i) r_i^{(\alpha)}.$$

So we can write $b_n^{s/2}(C_4) = (I) + (II) + (III)$, where

$$(I) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_i \delta'(x_i)\hat{H}^{-1}(x_i)\delta(x_i)h^2(x_i),$$

$$(II) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_i r_i^{(\beta)\prime}\hat{H}^{-1}(x_i)\delta(x_i)h(x_i), \quad (III) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_i r_i^{(\gamma)\prime}\hat{H}^{-1}(x_i)r_i^{(\alpha)}.$$

Therefore, using the facts that: $(i)$ $r_i^{(\alpha)}$ and $r_i^{(\beta)}$ are asymptotically negligible element by element, $(ii)$ the components of $\delta(x_i)$ and $r_i^{(\gamma)}$ are $O_p(1)$, and $(iii)$ $\hat{H}^{-1}(x_i)$ is element by element uniformly bounded in probability on $S_*$, it is easy to see that $(II), (III) = o_p(1)$. So it only remains to look at $(I)$. Write $(I) = (I)_a + (I)_b$, where

$$(I)_a = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_i \delta'(x_i)H^{-1}(x_i)\delta(x_i)h^2(x_i) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_i \delta'(x_i)V^{-1}(x_i)\delta(x_i),$$

$$(I)_b = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_i \delta'(x_i)\{\hat{H}^{-1}(x_i) - H^{-1}(x_i)\}\delta(x_i)h^2(x_i).$$

Then by the weak law of large numbers, $(I)_a = \mu + o_p(1)$. From Lemma B.10 we know that $\hat{H}^{-1}(x_i) \xrightarrow{p} H^{-1}(x_i)$ element by element under our choice of $b_n$. Hence we get $(I)_b = o_p(1)$. Therefore, $(I) = \mu + o_p(1)$ which implies that $b_n^{s/2}(C_4) = \mu + o_p(1)$. Combining the results for $(C_1)$–$(C_4)$ we obtain the desired conclusion. □

## Appendix B. Proofs of Auxiliary Results

**Lemma B.1.** *Let* $b_n = n^{-\alpha}$ *for* $0 < \alpha < \frac{1}{s}(1 - \frac{2}{m})$ *and* $c_n \stackrel{\text{def}}{=} \sqrt{\frac{\log n}{nb_n^s}}$. *Then under* $\mathrm{H}_0$,

$$\max_{i \in I_*} \|r_i^{(1)}\| = n^{1+1/m}O_p(c_n^2) \text{ and } \max_{i \in I_*, 1 \le j \le n} |\lambda_i' z_j| = n^{1+1/m}O_p(c_n).$$

**Proof of Lemma B.1** $w_{ij} \ge 0$ from Assumption 3.4, and $n + \lambda_i' z_j \ge 0$ holds because the estimated probabilities are nonnegative. Hence

$$\|r_i^{(1)}\| \le \sum_{j=1}^{n} \frac{w_{ij}(\lambda_i' z_j)^2}{n + \lambda_i' z_j}\|z_j\| \le z_n^* \sum_{j=1}^{n} \frac{w_{ij}(\lambda_i' z_j)^2}{n + \lambda_i' z_j},$$

where $z_n^* = \max\{\|z_1\|, \ldots, \|z_n\|\}$. Furthermore, multiplying both sides of (2.6) by $\lambda_i'$ we get

$$0 = \sum_{j=1}^{n} \frac{w_{ij}\lambda_i' z_j}{n + \lambda_i' z_j} = \sum_{j=1}^{n} \frac{w_{ij}}{n}\{\lambda_i' z_j - \frac{(\lambda_i' z_j)^2}{n + \lambda_i' z_j}\}; \quad \text{i.e.}$$

$$\sum_{j=1}^{n} \frac{w_{ij}(\lambda_i' z_j)^2}{n + \lambda_i' z_j} = \sum_{j=1}^{n} w_{ij}\lambda_i' z_j. \tag{B.1}$$

Plugging this result in the above inequality we have

$$\|r_i^{(1)}\| \leq z_n^* \sum_{j=1}^{n} w_{ij} \lambda_i' z_j.$$

Now use Lemma C.3 to assume w.l.o.g that $n$ is large enough so that $z_n^* < n^{1/m}$ holds almost surely. Thus the previous inequality reduces to

$$\|r_i^{(1)}\| \leq n^{1/m} \sum_{j=1}^{n} w_{ij} \lambda_i' z_j \tag{B.2}$$

for large enough $n$. Notice that $\mathbb{E}\{\frac{1}{nb_n^s}\sum_{j=1}^n \mathcal{K}_{ij} z_j\} = 0$ when $\mathbb{E}(z_j|x_j) = 0$. Therefore, we can use Newey (1994, Lemma B.1, Page 250) to show

$$\sup_{x_i \in S_*} |\frac{1}{nb_n^s} \sum_{j=1}^{n} \mathcal{K}_{ij} z_j| \overset{\text{elt. by elt.}}{=} O_p(c_n),$$

$$\sup_{x_i \in S_*} |\frac{1}{nb_n^s} \sum_{j=1}^{n} \mathcal{K}_{ij} - \mathbb{E}\{\frac{1}{nb_n^s} \sum_{j=1}^{n} \mathcal{K}_{ij}\}| = O_p(c_n),$$

and $c_n \downarrow 0$ on letting $b_n = n^{-\alpha}$ for $0 < \alpha < \frac{1}{s}(1 - \frac{2}{m})$. But since $h$ is bounded away from zero on $S_*$, which means that for large enough $n$ the term $\mathbb{E}\{\frac{1}{nb_n^s}\sum_{j=1}^n \mathcal{K}_{ij}\}$ is bounded away from zero on $S_*$, we can use Lemma C.2 to see that

$$\sup_{x_i \in S_*} |\sum_{j=1}^{n} w_{ij} z_j| \overset{\text{elt. by elt.}}{=} O_p(c_n).$$

Hence if we can determine a bound for the $\lambda_i$'s, we can bound $r_i^{(1)}$. So let us obtain the bound for $\lambda_i$. First, w.l.o.g let $\lambda_i = \rho_i \xi_i$, where $\rho_i \geq 0$ and $\xi_i \in \mathbb{S}$; i.e. $\|\xi_i\| = 1$. Therefore, (B.2) reduces to

$$\|r_i^{(1)}\| \leq n^{1/m} \rho_i O_p(c_n). \tag{B.3}$$

Next, use the fact that

$$0 \leq n + \lambda_i' z_j \leq n + \rho_i \|z_j\| \leq n + \rho_i z_n^* \leq n + \rho_i n^{1/m}$$

to see that $\frac{1}{n+\rho_i n^{1/m}} \leq \frac{1}{n+\lambda_i' z_j}$. But this implies that (B.1) becomes

$$\frac{\rho_i}{n + \rho_i n^{1/m}} \sum_{j=1}^{n} w_{ij} (\xi_i' z_j)^2 \leq \sum_{j=1}^{n} w_{ij} \xi_i' z_j;$$

i.e. using the nonnegativity of $\mathcal{K}_{ij}$ we can write (uniformly in $i \in I_*$)

$$\frac{\rho_i}{n + \rho_i n^{1/m}} \leq \frac{\sum_{j=1}^n w_{ij} \xi_i' z_j}{\xi_i' \hat{V}(x_i) \xi_i} = \frac{O_p(c_n)}{\xi_i' V(x_i) \xi_i + o_p(1)} = O_p(c_n),$$

where the last equality follows from the fact that $\hat{V}(x_i)$ converges in probability to $V(x_i)$ uniformly on $S_*$, and that $\sup_{x_i \in S_*} \xi_i' V(x_i)\xi_i < \infty$. Therefore, solving for $\rho_i$ we have

$$\rho_i = \|\lambda_i\| = \frac{n\, O_p(c_n)}{1 - n^{1/m} O_p(c_n)} = n O_p(c_n) \quad \text{(uniformly in } i \in I_*),$$

because $n^{1/m} c_n \downarrow 0$ under our choice of $b_n$. Substituting this in (B.3), we get

$$\|r_i^{(1)}\| = n^{1+1/m} O_p(c_n^2) \quad \text{(uniformly in } i \in I_*).$$

Moreover, since $|\lambda_i' z_j| \leq \|\lambda_i\| \|z_j\| \leq \rho_i z_n^*$, we also have

$$|\lambda_i' z_j| = n^{1+1/m} O_p(c_n) \quad \text{(uniformly in } i \in I_* \text{ and } 1 \leq j \leq n). \quad \square$$

**Lemma B.2.** $b_n^{s/2}\{\mathrm{T}_{n,5} - \mathrm{T}_{n,5}^*\} = o_p(1)$ *if* $b_n = n^{-\alpha}$ *and* $0 < \alpha < \frac{1}{s}(1 - \frac{2}{m})$.

**Proof of Lemma B.2**. Throughout this proof let

$$\mathcal{U} = \{\gamma \in \mathbb{R} : |\gamma| \leq 1\} \quad \text{and} \quad \Gamma_n(x_i) = \hat{H}^{-1}(x_i) - H^{-1}(x_i).$$

Changing the order of summation write

$$\begin{aligned} b_n^{s/2}\{\mathrm{T}_{n,5} - \mathrm{T}_{n,5}^*\} &= \frac{1}{n^2 b_n^{3s/2}} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \sum_{t=1, t\neq j\neq i}^{n} \mathbb{I}_i \mathcal{K}_{ij} z_j' \Gamma_n(x_i) z_t \mathcal{K}_{it} \\ &= \frac{1}{n^2 b_n^{3s/2}} \sum_{t=1}^{n} \sum_{j=1, j\neq t}^{n} \sum_{i=1, i\neq j\neq t}^{n} \mathbb{I}_i \mathcal{K}_{ij} z_j' \Gamma_n(x_i) z_t \mathcal{K}_{it} \\ &= \frac{1}{n^2 b_n^{3s/2}} \sum_{t=1}^{n} \sum_{j=1, j\neq t}^{n} z_j' B_{jt}(\Gamma_n) z_t, \end{aligned}$$

where $B_{jt}(\Gamma_n) = \sum_{i=1, i\neq j\neq t}^{n} \mathbb{I}_i \mathcal{K}_{ij} \Gamma_n(x_i) \mathcal{K}_{it}$. Note that since

$$z_j' B_{jt}(\Gamma_n) z_t = \sum_{l=1}^{d} \sum_{v=1}^{d} z_j^{(l)} B_{jt}(\Gamma_n^{(lv)}) z_t^{(v)},$$

it suffices to show that

$$A(\Gamma_n^{(lv)}) = \sum_{t=1}^{n} \sum_{j=1, j\neq t}^{n} z_j^{(l)} B_{jt}(\Gamma_n^{(lv)}) z_t^{(v)} = o_p(n^2 b_n^{3s/2}).$$

Before we present the details showing $A(\Gamma_n^{(lv)}) = o_p(n^2 b_n^{3s/2})$, we need two useful facts. The first one follows from Lemma B.10; namely,

$$\sup_{x_i \in S_*} |\Gamma_n^{(lv)}(x_i)| = O_p(\tau_n), \quad \tau_n \stackrel{\text{def}}{=} \sqrt{\frac{\log n}{n b_n^s}} + b_n^2,$$

and $\tau_n \downarrow 0$ on choosing $b_n = n^{-\alpha}$ for $0 < \alpha < \frac{1}{s}(1 - \frac{2}{m})$. Hereafter, for $\gamma \in \mathcal{U}$ we will also use the following notation:

$$B^*_{jt}(\gamma) = \sum_{i=1, i\neq j\neq t}^{n} \mathbb{I}_i \mathcal{K}_{ij}\gamma\mathcal{K}_{it}, \qquad A^*(\gamma) = \sum_{t=1}^{n} \sum_{j=1, j\neq t}^{n} z_j^{(l)} B^*_{jt}(\gamma) z_t^{(v)}.$$

The second useful fact, which is easily verified, is that the random variables $z_j^{(l)} B^*_{jt}(\gamma) z_t^{(v)}$ and $z_k^{(l)} B^*_{kt}(\gamma) z_t^{(v)}$ are uncorrelated for $j \neq k \neq t$ when $\gamma \in \mathcal{U}$. See, for example, Remark B.1 in the proof of Lemma B.3. Let us continue with the proof. So pick any $\epsilon > 0$, and let $M_\epsilon$ denote a positive number which may depend upon $\epsilon$. Observe that

$$\begin{aligned}
\text{(B.4)} \quad & \Pr\{|A(\tau_n^{-1/2}\Gamma_n^{(lv)})| > M_\epsilon\} \\
&= \Pr\{|A(\tau_n^{-1/2}\Gamma_n^{(lv)})| > M_\epsilon, \sup_{x_i \in S_*} \tau_n^{-1/2} |\Gamma_n^{(lv)}(x_i)| \leq 1\} \\
&\quad + \Pr\{|A(\tau_n^{-1/2}\Gamma_n^{(lv)})| > M_\epsilon, \sup_{x_i \in S_*} \tau_n^{-1/2} |\Gamma_n^{(lv)}(x_i)| > 1\}.
\end{aligned}$$

Since $\sup_{x_i \in S_*} \tau_n^{-1/2} |\Gamma_n^{(lv)}(x_i)| \leq 1$ implies that

$$\tau_n^{-1/2}\Gamma_n^{(lv)}(x_1)\mathbb{I}_1 \in \mathcal{U}, \ldots, \tau_n^{-1/2}\Gamma_n^{(lv)}(x_n)\mathbb{I}_n \in \mathcal{U},$$

the first term on the RHS of (B.4) is majorized by $\Pr\{\sup_{\gamma\in\mathcal{U}} |A^*(\gamma)| > M_\epsilon\}$. Moreover, from the first useful fact, the second term on the RHS of (B.4) is $o(1)$. Therefore, we have

$$\Pr\{|A(\tau_n^{-1/2}\Gamma_n^{(lv)})| > M_\epsilon\} \leq \Pr\{\sup_{\gamma\in\mathcal{U}} |A^*(\gamma)| > M_\epsilon\} + o(1).$$

Because $A^*(\gamma)$ is linearly homogeneous in $\gamma$, $\sup_{\gamma\in\mathcal{U}} |A^*(\gamma)| \leq |A^*(1)|$. Hence by Chebychev, the previous inequality reduces to

$$\text{(B.5)} \qquad \Pr\{|A(\tau_n^{-1/2}\Gamma_n^{(lv)})| > M_\epsilon\} \leq \mathbb{E}|A^*(1)|/M_\epsilon + o(1).$$

Now using (D.2) along with the second useful fact, we can see that

$$\{\mathbb{E}|A^*(1)|\}^2 \leq \mathbb{E}\{A^*(1)\}^2 = 2\sum_{t=1}^{n} \sum_{j=1, j\neq t}^{n} \mathbb{E}\{z_j^{(l)} B^*_{jt}(1) z_t^{(v)}\}^2.$$

Because

$$\{B^*_{jt}(1)\}^2 = \sum_{i=1, i\neq j\neq t}^{n} \mathbb{I}_i \mathcal{K}^2_{ij}\mathcal{K}^2_{it} + \sum_{i=1, i\neq j\neq t}^{n} \sum_{u=1, u\neq i\neq j\neq t}^{n} \mathbb{I}_i \mathcal{K}_{ij}\mathcal{K}_{it}\mathcal{K}_{uj}\mathcal{K}_{ut},$$

we have

$$\begin{aligned}
\mathbb{E}\{z_j^{(l)} B^*_{jt}(1) z_t^{(v)}\}^2 &= \sum_{i=1, i\neq j\neq t}^{n} \mathbb{E}\{\mathbb{I}_i [z_j^{(l)}]^2 \, [z_t^{(v)}]^2 \mathcal{K}^2_{ij}\mathcal{K}^2_{it}\} \\
&\quad + \sum_{i=1, i\neq j\neq t}^{n} \sum_{u=1, u\neq i\neq j\neq t}^{n} \mathbb{E}\{\mathbb{I}_i [z_j^{(l)}]^2 \, [z_t^{(v)}]^2 \mathcal{K}_{ij}\mathcal{K}_{it}\mathcal{K}_{uj}\mathcal{K}_{ut}\}.
\end{aligned}$$

It is straightforward (the details being very similar to the method used in the proof of Lemma B.4) to see that

$$\mathbb{E}\{\mathbb{I}_i[z_j^{(l)}]^2[z_t^{(v)}]^2\mathcal{K}_{ij}^2\mathcal{K}_{it}^2\} = O(b_n^{2s}).$$

Similarly, we can show that

$$\mathbb{E}\{\mathbb{I}_i[z_j^{(l)}]^2\,[z_t^{(v)}]^2\mathcal{K}_{ij}\mathcal{K}_{it}\mathcal{K}_{uj}\mathcal{K}_{ut}\} = O(b_n^{3s}).$$

Therefore, using the above results we have

$$\mathbb{E}\{z_j^{(l)}B_{jt}^*(1)z_t^{(v)}\}^2 = O(n^2b_n^{3s}).$$

Hence $\mathbb{E}\{A^*(1)\}^2 = O(n^4b_n^{3s})$, which implies that $\mathbb{E}|A^*(1)| = O(n^2b_n^{3s/2})$. Thus (B.5) reduces to

$$\Pr\{|A(\tau_n^{-1/2}\Gamma_n^{(lv)})| > M_\epsilon\} \le O(n^2b_n^{3s/2})/M_\epsilon + o(1).$$

So choosing $M_\epsilon$ appropriately, we obtain

$$A(\tau_n^{-1/2}\Gamma_n^{(lv)}) = O_p(n^2b_n^{3s/2}).$$

But since $\tau_n \downarrow 0$, we conclude that $A(\Gamma_n^{(lv)}) = o_p(n^2b_n^{3s/2})$. □

**Lemma B.3.** *$W_{tjn}$ is clean; i.e. $\mathbb{E}(W_{tjn}|x_t, z_t) = 0$ a.s. for all $1 \le t, j \le n$.*
**Proof of Lemma B.3** Since

$$\mathbb{E}(W_{tjn}|x_t, z_t) = 2\mathbb{E}(z_t'A_{tjn}z_j|x_t, z_t) = 2z_t'\mathbb{E}(A_{tjn}z_j|x_t, z_t),$$

use iterated expectations to see that

$$\mathbb{E}(A_{tjn}z_j|x_t, z_t) = \mathbb{E}\{\mathbb{E}(A_{tjn}z_j|x_t, z_t)|\mathcal{X}, z_t\} = \mathbb{E}\{A_{tjn}\mathbb{E}(z_j|\mathcal{X}, z_t)|x_t, z_t\}.$$

But as we are dealing with independent observations, under $\mathrm{H}_0$

$$\mathbb{E}(z_j|\mathcal{X}, z_t) = \mathbb{E}(z_j|x_j) = 0;$$

i.e. $\mathbb{E}(W_{tjn}|x_t, z_t) = 0$. Similarly, we can show that $\mathbb{E}(W_{tjn}|x_j, z_j) = 0$.
**Remark B.1.** We can use a similar approach to show that $W_{tjn}$ and $W_{tkn}$ are uncorrelated for $t \ne j \ne k$. To see this, note that

$$\begin{aligned}\mathbb{E}\{W_{tjn}W_{tkn}\} &= 4\mathbb{E}\{z_t'A_{tjn}z_jz_k'A_{tkn}z_t\}\\ &= 4\mathbb{E}\{z_t'A_{tjn}z_j\mathbb{E}(z_k'|\mathcal{X}, z_t, z_j)A_{tkn}z_t\}\\ &= 4\mathbb{E}\{z_t'A_{tjn}z_j\mathbb{E}(z_k'|x_k)A_{tkn}z_t\} = 0;\end{aligned}$$

i.e. $W_{tjn}$ and $W_{tkn}$ are uncorrelated if a free[8] index is present. □

**Lemma B.4.** $\mathbb{E}\{z_t'A_{tjn}z_j\}^2 = (n-2)(n-3)2db_n^{3s}\mathcal{K}^{**}vol(S_*)\{1 + o(1)\}$.

[8]An index is "free" if it occurs only once (de Jong 1987, Lemma 2.1, Page 263).

**Proof of Lemma B.4**. Using iterated expectations and the independence of observations, we can write

$$\begin{aligned}\mathbb{E}\{z_t' A_{tjn} z_j\}^2 &= \mathbb{E}\{z_t' A_{tjn} z_j z_j' A_{tjn} z_t\} = \mathbb{E}\{z_t' A_{tjn} \mathbb{E}[z_j z_j' | \mathcal{X}, z_t] A_{tjn} z_t\} \\ &= \mathbb{E}\{z_t' A_{tjn} \mathbb{E}[z_j z_j' | x_j] A_{tjn} z_t\} = \mathbb{E}\{z_t' A_{tjn} V(x_j) A_{tjn} z_t\} \\ &= \mathbb{E}\, tr\, \{z_t' A_{tjn} V(x_j) A_{tjn} z_t\} = tr\, \mathbb{E}\{A_{tjn} V(x_j) A_{tjn} z_t z_t'\} \\ &= tr\, \mathbb{E}\{A_{tjn} V(x_j) A_{tjn} \mathbb{E}[z_t z_t' | \mathcal{X}]\} \\ &= tr\, \mathbb{E}\{A_{tjn} V(x_j) A_{tjn} \mathbb{E}[z_t z_t' | x_t]\} \\ &= tr\, \mathbb{E}\{A_{tjn} V(x_j) A_{tjn} V(x_t)\}.\end{aligned}$$

But as $A_{tjn}V(x_j)A_{tjn}V(x_t)$ is equal to

$$\{\sum_{i=1, i\neq j\neq t}^{n} \mathbb{I}_i \mathcal{K}_{ij} H^{-1}(x_i) V(x_j) \mathcal{K}_{it}\} \times \{\sum_{u=1, u\neq j\neq t}^{n} \mathbb{I}_u \mathcal{K}_{uj} H^{-1}(x_u) V(x_t) \mathcal{K}_{ut}\},$$

we get that

$$\begin{aligned}\mathbb{E}\{z_t' A_{tjn} z_j\}^2 &= \sum_{i=1, i\neq j\neq t}^{n} tr\, P_1 + \sum_{i=1, i\neq j\neq t}^{n} \sum_{u=1, u\neq i\neq j\neq t}^{n} tr\, P_2, \quad \text{where} \\ P_1 &= \mathbb{E}\{\frac{\mathbb{I}_i \mathcal{K}_{ij}^2 \mathcal{K}_{it}^2 V^{-1}(x_i) V(x_j) V^{-1}(x_i) V(x_t)}{h^4(x_i)}\}, \\ P_2 &= \mathbb{E}\{\frac{\mathbb{I}_i \mathbb{I}_u \mathcal{K}_{ij} \mathcal{K}_{it} \mathcal{K}_{uj} \mathcal{K}_{ut} V^{-1}(x_i) V(x_j) V^{-1}(x_u) V(x_t)}{h^2(x_i)\, h^2(x_u)}\}.\end{aligned}$$

Let us look at $P_1$ and $P_2$ one by one. By iterated expectations and the independence of observations we can write

$$P_1 = \mathbb{E}\{\frac{\mathbb{I}_i \mathcal{K}_{ij}^2 V^{-1}(x_i) V(x_j) V^{-1}(x_i)}{h^4(x_i)} \mathbb{E}[\mathcal{K}_{it}^2 V(x_t) | x_i]\}.$$

But using Assumption 3.3($ii$) and the fact that $x_i \in S_*$, we can show

$$\mathbb{E}[\mathcal{K}_{it}^2 V(x_t) | x_i] = b_n^s \mathfrak{R}(\mathcal{K}) V(x_i) h(x_i) + b_n^{s+2} R^{(3)}(x_i), \tag{B.6}$$

where $\sup_{x_i \in S_*} \|R^{(3)}(x_i)\| < \infty$. Hence substituting (B.6) in the expression for $P_1$, we have

$$\begin{aligned}P_1 = b_n^s \mathfrak{R}(\mathcal{K}) \mathbb{E}\{&\frac{\mathbb{I}_i \mathcal{K}_{ij}^2 V^{-1}(x_i) V(x_j)}{h^3(x_i)}\} \\ &+ b_n^{s+2} \mathbb{E}\{\frac{\mathbb{I}_i \mathcal{K}_{ij}^2 V^{-1}(x_i) V(x_j) V^{-1}(x_i) R^{(3)}(x_i)}{h^4(x_i)}\}.\end{aligned}$$

But note that

$$\begin{aligned}\mathbb{E}\{\frac{\mathbb{I}_i\mathcal{K}_{ij}^2V^{-1}(x_i)V(x_j)}{h^3(x_i)}\} &= \mathbb{E}\{\frac{\mathbb{I}_iV^{-1}(x_i)}{h^3(x_i)}\mathbb{E}[\mathcal{K}_{ij}^2V(x_j)|x_i]\}\\ &= b_n^s\mathfrak{R}(\mathcal{K})\mathbb{E}\{\frac{\mathbb{I}_i}{h^2(x_i)}\}I_{d\times d} + b_n^{s+2}\mathbb{E}\{\frac{\mathbb{I}_iV^{-1}(x_i)R^{(4)}(x_i)}{h^3(x_i)}\},\end{aligned}$$

where $\|R^{(4)}(x_i)\| < \infty$ uniformly in $x_i \in S_*$, and the last equality follows from a result similar to the one obtained in (B.6). Hence using the facts: $(i)$ $\sup_{x_i\in S_*}\|V^{-1}(x_i)\| < \infty$, $(ii)$ $R^{(4)}(x_i)$ is element by element uniformly bounded on $S_*$, and $(iii)$ $h$ is bounded away from zero on $S_*$, we have

$$\begin{aligned}b_n^s\mathfrak{R}(\mathcal{K})tr\,\mathbb{E}\{\frac{\mathbb{I}_i\mathcal{K}_{ij}^2V^{-1}(x_i)V(x_j)}{h^3(x_i)}\} &= db_n^{2s}\mathfrak{R}^2(\mathcal{K})\mathbb{E}\{\frac{\mathbb{I}_i}{h^2(x_i)}\}\\ &\quad + \mathfrak{R}(\mathcal{K})b_n^{2s+2}tr\,\mathbb{E}\{\frac{\mathbb{I}_iV^{-1}(x_i)R^{(4)}(x_i)}{h^3(x_i)}\}\\ &= db_n^{2s}\mathfrak{R}^2(\mathcal{K})\mathbb{E}\{\frac{\mathbb{I}_i}{h^2(x_i)}\} + O(b_n^{2s+2}).\end{aligned}$$

Similarly, we can show that

$$\begin{aligned}&b_n^{s+2}tr\,\mathbb{E}\{\frac{\mathbb{I}_i\mathcal{K}_{ij}^2V^{-1}(x_i)V(x_j)V^{-1}(x_i)R^{(3)}(x_i)}{h^3(x_i)}\}\\ &\quad= b_n^{s+2}tr\,\mathbb{E}\{\frac{\mathbb{I}_i}{h^3(x_i)}V^{-1}(x_i)\mathbb{E}[\mathcal{K}_{ij}^2V(x_j)|x_i]V^{-1}(x_i)R^{(3)}(x_i)\}\\ &\quad= b_n^{2s+2}\mathfrak{R}(\mathcal{K})tr\,\mathbb{E}\{\frac{\mathbb{I}_iV^{-1}(x_i)R^{(3)}(x_i)}{h^2(x_i)}\}\\ &\qquad\quad + b_n^{2s+4}tr\,\mathbb{E}\{\frac{\mathbb{I}_iV^{-1}(x_i)R^{(5)}(x_i)V^{-1}(x_i)R^{(3)}(x_i)}{h^3(x_i)}\}\\ &\quad= O(b_n^{2s+2}) + O(b_n^{2s+4}) = O(b_n^{2s+2}),\end{aligned}$$

where the last line follows from the observation that $R^{(3)}(x_i)$, $R^{(5)}(x_i)$, and $V^{-1}(x_i)$ are element by element uniformly bounded on $S_*$ and $h$ is bounded away from zero on $S_*$. Therefore, we have

$$tr\,P_1 = db_n^{2s}\mathfrak{R}^2(\mathcal{K})\mathbb{E}\{\frac{\mathbb{I}_1}{h^2(x_1)}\}[1 + O(b_n^2)].$$

Now let us look at $P_2$. Once again, using iterated expectations and the independence of observations, we can write

$$P_2 = \mathbb{E}\{\frac{\mathbb{I}_i\mathbb{I}_u\mathcal{K}_{ij}\mathcal{K}_{uj}V^{-1}(x_i)V(x_j)V^{-1}(x_u)}{h^2(x_i)h^2(x_u)}\mathbb{E}[\mathcal{K}_{it}\mathcal{K}_{ut}V(x_t)|x_i,x_u]\}.$$

Next, using Assumption 3.3$(ii)$ and keeping in mind that $x_i, x_u \in S_*$, we can show that

$$\mathbb{E}\{\mathcal{K}_{it}\mathcal{K}_{ut}V(x_t)|x_i,x_u\} = b_n^s\mathcal{K}_{iu}^*V(x_u)h(x_u) + b_n^{s+1}\tilde{R}^{(1)}(x_i,x_u),$$

where $\mathcal{K}^*_{iu} = \mathcal{K}^*(\frac{x_i - x_u}{b_n})$ and $\sup_{x_i, x_u \in S_*} \|\tilde{R}^{(1)}(x_i, x_u)\| < \infty$. Using this result, the expression for $P_2$ reduces to

$$P_2 = b_n^s \mathbb{E}\{\frac{\mathbb{I}_i \mathbb{I}_u \mathcal{K}_{ij} \mathcal{K}_{uj} \mathcal{K}^*_{iu} V^{-1}(x_i) V(x_j)}{h^2(x_i) h(x_u)}\} \\ + b_n^{s+1} \mathbb{E}\{\frac{\mathbb{I}_i \mathbb{I}_u \mathcal{K}_{ij} \mathcal{K}_{uj} V^{-1}(x_i) V(x_j) V^{-1}(x_u) \tilde{R}^{(1)}(x_i, x_u)}{h^2(x_i) h^2(x_u)}\}.$$

But once again we can show that

$$\mathbb{E}\{\frac{\mathbb{I}_i \mathbb{I}_u \mathcal{K}_{ij} \mathcal{K}_{uj} \mathcal{K}^*_{iu} V^{-1}(x_i) V(x_j)}{h^2(x_i) h(x_u)}\} \\ = \mathbb{E}\{\frac{\mathbb{I}_i \mathbb{I}_u \mathcal{K}^*_{iu} V^{-1}(x_i)}{h^2(x_i) h(x_u)} \mathbb{E}[\mathcal{K}_{ij} \mathcal{K}_{uj} V(x_j) | x_i, x_u]\} \\ = \mathbb{E}\{\frac{\mathbb{I}_i \mathbb{I}_u \mathcal{K}^*_{iu} V^{-1}(x_i)}{h^2(x_i) h(x_u)} [b_n^s \mathcal{K}^*_{ui} V(x_i) h(x_i) + b_n^{s+1} \tilde{R}^{(2)}(x_u, x_i)]\} \\ = b_n^s \mathbb{E}\{\frac{\mathbb{I}_i \mathbb{I}_u \mathcal{K}^*_{iu} \mathcal{K}^*_{ui}}{h(x_i) h(x_u)}\} I_{d \times d} + b_n^{s+1} \mathbb{E}\{\frac{\mathbb{I}_i \mathbb{I}_u \mathcal{K}^*_{iu} V^{-1}(x_i) \tilde{R}^{(2)}(x_u, x_i)}{h^2(x_i) h(x_u)}\},$$

for some $\tilde{R}^{(2)}(x_u, x_i)$ such that $\sup_{x_u, x_i \in S_*} \|\tilde{R}^{(2)}(x_u, x_i)\| < \infty$. Furthermore, after some tedious algebra we can also show the following results:

$$tr\, \mathbb{E}\{\frac{\mathbb{I}_i \mathbb{I}_u \mathcal{K}^*_{iu} \mathcal{K}^*_{ui}}{h(x_i) h(x_u)}\} I_{d \times d} = d b_n^s \mathcal{K}^{**} vol(S_*),$$

$$tr\, \mathbb{E}\{\frac{\mathbb{I}_i \mathbb{I}_u \mathcal{K}^*_{iu} V^{-1}(x_i) \tilde{R}^{(2)}(x_u, x_i)}{h^2(x_i) h(x_u)}\} = O(b_n^s),$$

$$tr\, \mathbb{E}\{\frac{\mathbb{I}_i \mathbb{I}_u \mathcal{K}_{ij} \mathcal{K}_{uj} V^{-1}(x_i) V(x_j) V^{-1}(x_u) \tilde{R}^{(1)}(x_i, x_u)}{h^2(x_i) h^2(x_u)}\} = O(b_n^{2s}) + O(b_n^{2s+1}).$$

Therefore, combining these results we get

$$tr\, P_2 = d b_n^{3s} \mathcal{K}^{**} vol(S_*) \{1 + O(b_n)\}.$$

The desired result follows. $\square$

**Lemma B.5.** $G_I = O(n^{4+\frac{8}{m}} b_n^{2s})$.

**Proof of Lemma B.5**. Since

$$G_I = \sum_{t=1}^{n-1} \sum_{j=t+1}^{n} \mathbb{E} W_{tjn}^4 = 16 \sum_{t=1}^{n-1} \sum_{j=t+1}^{n} \mathbb{E}\{z_t' A_{tjn} z_j\}^4,$$

use the summation identity in (D.4) to see that

$$\mathbb{E}\{z_t' A_{tjn} z_j\}^4 = \sum_{i=1, i\neq j\neq t}^{n} \mathbb{E}\{\mathbb{I}_i \mathcal{K}_{ij} z_t' H^{-1}(x_i) z_j \mathcal{K}_{it}\}^4$$

$$+3 \sum_{i=1, i\neq j\neq t}^{n} \sum_{k=1, k\neq i\neq j\neq t}^{n} \mathbb{E}\{\mathbb{I}_i \mathcal{K}_{ij} z_t' H^{-1}(x_i) z_j \mathcal{K}_{it}\}^2 \{\mathbb{I}_k \mathcal{K}_{kj} z_t' H^{-1}(x_k) z_j \mathcal{K}_{kt}\}^2.$$

Now $|z_t' H^{-1}(x_i) z_j| \leq \frac{\|z_t\| \, \|z_j\| \, \|V^{-1}(x_i)\|}{h^2(x_i)}$. Therefore, by Assumption 3.3($iii$) and the result of Lemma C.3, $|z_t' H^{-1}(x_i) z_j| \leq c n^{\frac{2}{m}} h^{-2}(x_i)$ holds almost surely for large enough $n$. Thus

$$\mathbb{E}\{\mathbb{I}_i \mathcal{K}_{ij} z_t' H^{-1}(x_i) z_j \mathcal{K}_{it}\}^4 \leq c n^{\frac{8}{m}} \mathbb{E}\{\frac{\mathbb{I}_i \mathcal{K}_{ij}^4 \mathcal{K}_{it}^4}{h^2(x_i)}\}$$

also holds for large enough $n$. Next, a little algebra reveals that

$$\mathbb{E}\{\frac{\mathbb{I}_i \mathcal{K}_{ij}^4 \mathcal{K}_{it}^4}{h^2(x_i)}\} = O(b_n^{2s}).$$

But this means that

$$\mathbb{E}\{\mathbb{I}_i \mathcal{K}_{ij} z_t' H^{-1}(x_i) z_j \mathcal{K}_{it}\}^4 = O(n^{\frac{8}{m}} b_n^{2s}).$$

The result now follows from another application of Cauchy-Schwarz. □

**Lemma B.6.** $G_{II} = O(n^{5+\frac{8}{m}} b_n^{2s})$.

**Proof of Lemma B.6**. Follows by using the Cauchy-Schwarz inequality and the result that $\mathbb{E}W_{tjn}^4 = O(n^{2+\frac{8}{m}} b_n^{2s})$ (see proof of Lemma B.5). □

**Lemma B.7.** $G_{IV} = O(n^{6+\frac{8}{m}} b_n^{2s})$.

**Proof of Lemma B.7**. Same as the proof of Lemma B.6. □

**Lemma B.8.** $b_n^{s/2}(C_2)_a = o_p(1)$.

**Proof of Lemma B.8**. Change the order of summation to write

$$(C_2)_a = \frac{\epsilon_n}{n^2 b_n^{2s}} \sum_{t=1}^{n} \sum_{j=1, j\neq t}^{n} \delta'(x_t) A_{tjn} \tilde{z}_j, \quad A_{tjn} = \sum_{i=1, i\neq j\neq t}^{n} \mathbb{I}_i \mathcal{K}_{ij} H^{-1}(x_i) \mathcal{K}_{it}.$$

For notational convenience let $b_{tj} = \delta'(x_t) A_{tjn} \tilde{z}_j$ and note that $b_{tj} \neq b_{jt}$, $\mathbb{E}\{b_{tj} b_{jt}\} = 0$ if $t \neq j$, and $\mathbb{E}\{b_{tj} b_{mu}\} = 0$ if $u \neq j$. Then using the summation identity in (D.2), we have

$$\mathbb{E}\{b_n^{s/2}(C_2)_a\}^2 = \frac{\epsilon_n^2}{n^4 b_n^{3s}} \{\sum_{t=1}^{n} \sum_{j=1, j\neq t}^{n} \mathbb{E} b_{tj}^2 + \sum_{t=1}^{n} \sum_{j=1, j\neq t}^{n} \sum_{m=1, m\neq j\neq t}^{n} \mathbb{E} b_{tj} b_{mj}\}.$$

Recall that $\{\tilde{z}_i\}_{i=1}^n$ is a random sample and $\mathbb{E}(\tilde{z}_i|x_i) = 0$. Therefore, as in the proof of Lemma B.4, we can show that $\mathbb{E}b_{tj}^2 = O(n^2 b_n^{3s})$. So it remains to calculate $\mathbb{E}b_{tj}b_{mj}$. Using iterated expectations,

$$\begin{aligned}\mathbb{E}\{b_{tj}b_{mj}\} &= \mathbb{E}\{\delta'(x_t)A_{tjn}\mathbb{E}(\tilde{z}_j\tilde{z}_j'|x_j)A_{mjn}\delta(x_m)\} \\ &= \mathbb{E}\{\delta'(x_t)A_{tjn}\mathbb{E}(z_j z_j'|x_j)A_{mjn}\delta(x_m)\} \\ &= \mathbb{E}\{\delta'(x_t)A_{tjn}V(x_j)A_{mjn}\delta(x_m)\},\end{aligned}$$

where the second equality follows because $\mathbb{E}(\tilde{z}_j\tilde{z}_j'|x_j) = V(x_j)$. But as $A_{tjn}V(x_j)A_{mjn}$ is equal to

$$\{\sum_{i=1,i\neq j\neq t}^n \mathbb{I}_i\mathcal{K}_{it}\mathcal{K}_{ij}H^{-1}(x_i)\}\, V(x_j) \times \{\sum_{u=1,u\neq j\neq m}^n \mathbb{I}_u\mathcal{K}_{um}\mathcal{K}_{uj}H^{-1}(x_u)\},$$

we get that $\mathbb{E}\{b_{tj}b_{mj}\} = \mathbb{E}\{(IV)\times(V)\}$ where

$$\begin{aligned}(IV) &= \sum_{i=1,i\neq j\neq t}^n \mathbb{I}_i\mathcal{K}_{it}\mathcal{K}_{ij}\delta'(x_t)H^{-1}(x_i)V(x_j), \\ (V) &= \sum_{u=1,u\neq j\neq m}^n \mathbb{I}_u\mathcal{K}_{um}\mathcal{K}_{uj}H^{-1}(x_u)\delta(x_m).\end{aligned}$$

So we can write

$$\mathbb{E}\{b_{tj}b_{mj}\} = \sum_{i=1,i\neq j\neq t}^n (VI)_a + \sum_{i=1,i\neq j\neq t}^n \sum_{u=1,u\neq j\neq m}^n (VI)_b,$$

where

$$\begin{aligned}(VI)_a &= \mathbb{E}\{\mathbb{I}_i\mathcal{K}_{it}\mathcal{K}_{ij}^2\mathcal{K}_{im}\,\delta'(x_t)H^{-1}(x_i)V(x_j)H^{-1}(x_i)\delta(x_m)\} \\ (VI)_b &= \mathbb{E}\{\mathbb{I}_i\mathbb{I}_u\mathcal{K}_{it}\mathcal{K}_{ij}\mathcal{K}_{um}\mathcal{K}_{uj}\,\delta'(x_t)H^{-1}(x_i)V(x_j)H^{-1}(x_u)\delta(x_m)\}.\end{aligned}$$

We now sketch out the argument which shows that $(VI)_a = O(b_n^{3s})$ and $(VI)_b = O(b_n^{4s})$. First, using iterated expectations and the independence of observations, we can show that $(VI)_a$ is equal to

$$\mathbb{E}\{\mathbb{I}_i\mathbb{E}[\delta'(x_t)\mathcal{K}_{it}|x_i]\,H^{-1}(x_i)\,\mathbb{E}[V(x_j)\mathcal{K}_{ij}^2|x_i]\,H^{-1}(x_i)\,\mathbb{E}[\delta(x_m)\mathcal{K}_{im}|x_i]\}.$$

Then, keeping in mind the discussion after Assumption 5.1, we can show that (element by element) the following results hold uniformly for $x_i \in S_*$:

$$\mathbb{E}[\delta'(x_t)\mathcal{K}_{it}|x_i] = O(b_n^s),\ \mathbb{E}[V(x_j)\mathcal{K}_{ij}^2|x_i] = O(b_n^s),\ \mathbb{E}[\delta(x_m)\mathcal{K}_{im}|x_i] = O(b_n^s).$$

Therefore, since $H^{-1}(x_i)$ is element by element uniformly bounded on $S_*$, we get that $(VI)_a = O(b_n^{3s})$. Next we look at $(VI)_b$. Using iterated expectations once again, we can show that

$$\begin{aligned}(VI)_b = \mathbb{E}\{\mathbb{I}_i\mathbb{I}_u\mathbb{E}[\delta'(x_t)\mathcal{K}_{it}|x_i]\,H^{-1}(x_i)\,\mathbb{E}[V(x_j)\mathcal{K}_{ij}\mathcal{K}_{uj}|x_i,x_u]\,H^{-1}(x_u) \\ \times\,\mathbb{E}[\delta(x_m)\mathcal{K}_{um}|x_u]\}.\end{aligned}$$

But since the maps

$$\tilde{\mathcal{K}} : x_i \mapsto \sup_{u \in S_{\mathcal{K}}} \mathcal{K}(x_i + u),\ x_i \mapsto \sup_{u \in S_{\mathcal{K}}} h(x_i + b_n u),\ x_i \mapsto \sup_{u \in S_{\mathcal{K}}} \|V(x_i + b_n u)\|$$

are continuous on $S$, we can write

$$\mathbb{I}_i \mathbb{I}_u \mathbb{E}[V(x_j)\mathcal{K}_{ij}\mathcal{K}_{uj}|x_i, x_u] \overset{\text{elt. by elt.}}{\leq} b_n^s \mathbb{I}_i \mathbb{I}_u \tilde{\mathcal{K}}(\frac{x_i - x_u}{b_n}) M^*$$

where $M^*$ is a matrix of constants which does not depend upon $(x_i, x_u)$. Moreover, as pointed out earlier, we can also show that element by element

$$\mathbb{E}\{\delta'(x_t)\mathcal{K}_{it}|x_i\} = O(b_n^s), \quad \mathbb{E}\{\delta(x_m)\mathcal{K}_{um}|x_u\} = O(b_n^s)$$

hold uniformly in $x_i, x_u \in S_*$. Therefore, since $H^{-1}(x_i)$ and $H^{-1}(x_u)$ are element by element uniformly bounded on $S_*$, we get that

$$(VI)_b \leq c b_n^{3s} \mathbb{E}\{\mathbb{I}_i \mathbb{I}_u \tilde{\mathcal{K}}(\frac{x_i - x_u}{b_n})\}.$$

However, using the facts that $(i)$ $\tilde{K}$ is bounded on $[-2, 2]^s$, and $(ii)$ $x_i \mapsto \sup_{u \in S_{\mathcal{K}}} h(x_i + b_n u)$ and $x_i \mapsto h(x_i)$ are bounded on $S_*$, it is easy to show that $\mathbb{E}\{\mathbb{I}_i \mathbb{I}_u \tilde{\mathcal{K}}(\frac{x_i - x_u}{b_n})\} = O(b_n^s)$. Therefore, $(VI)_b = O(b_n^{4s})$, and we get

$$\mathbb{E}\{b_{tj} b_{mj}\} = O(n b_n^{3s}) + O(n^2 b_n^{4s}) = O(n^2 b_n^{4s})\{1 + O(\frac{1}{n b_n^s})\} = O(n^2 b_n^{4s}).$$

Combining the results for $\mathbb{E} b_{tj}^2$ and $\mathbb{E} b_{tj} b_{mj}$, it follows that

$$\begin{aligned}\mathbb{E}\{b_n^{s/2}(C_2)_a\}^2 &= \frac{\epsilon_n^2}{n^4 b_n^{3s}}\{O(n^4 b_n^{3s}) + O(n^5 b_n^{4s})\} \\ &= \frac{\epsilon_n^2}{n^4 b_n^{3s}} O(n^5 b_n^{4s}) = O(b_n^{s/2});\end{aligned}$$

i.e. $\mathbb{E}\{b_n^{s/2}(C_2)_a\}^2 = o(1)$. The desired conclusion follows. □

**Lemma B.9.** $b_n^{s/2}(C_2)_b = o_p(1)$.

**Proof of Lemma B.9**. Since the details here are very similar to the ones in the proof of Lemma B.2, we provide only the barest outline. So write

$$b_n^{s/2}(C_2)_b = \frac{\epsilon_n}{n^2 b_n^{3s/2}} \sum_{i=1}^n \sum_{j=1, j \neq i}^n \sum_{t=1, t \neq j \neq i}^n \mathbb{I}_i \mathcal{K}_{ij} \tilde{z}_j' \Gamma_n(x_i) \delta(x_t) \mathcal{K}_{it}.$$

Hence it suffices to show that $\epsilon_n A(\Gamma_n^{(lv)}) = o_p(n^2 b_n^{3s/2})$, where

$$A(\Gamma_n^{(lv)}) = \sum_{i=1}^n \sum_{j=1, j \neq i}^n \sum_{t=1, t \neq j \neq i}^n \mathbb{I}_i \mathcal{K}_{ij} \tilde{z}_j^{(l)} \Gamma_n^{(lv)}(x_i) \delta^{(v)}(x_t) \mathcal{K}_{it}.$$

But this can be shown almost exactly as in the proof of Lemma B.2. □

**Lemma B.10.** *Let* $\frac{n^{1-2/m}b_n^s}{\log n} \to \infty$ *and suppose that the assumptions in Newey (1994, Lemma B.3, Page 252) are satisfied. Then*

$$\sup_{x_i \in S_*} |\hat{H}^{-1}(x_i) - H^{-1}(x_i)| \overset{\text{elt. by elt.}}{=} O_p(\sqrt{\frac{\log n}{nb_n^s}} + b_n^2).$$

**Proof of Lemma B.10**. Define $\tau_n = \sqrt{\frac{\log n}{nb_n^s}} + b_n^2$. From Newey (1994, Lemma B.3, Page 252) we have

$$\sup_{x_i \in S_*} |\frac{1}{nb_n^s}\sum_{j=1}^{n} \mathcal{K}_{ij} z_j z_j' - V(x_i)h(x_i)| \overset{\text{elt. by elt.}}{=} O_p(\tau_n), \quad \text{and}$$

$$\sup_{x_i \in S_*} |\hat{h}(x_i) - h(x_i)| = O_p(\tau_n).$$

Since $h$ is bounded away from zero on $S_*$, we can use Lemma C.1 to show

$$\sup_{x_i \in S_*} |\hat{V}(x_i) - V(x_i)| \overset{\text{elt. by elt.}}{=} O_p(\tau_n).$$

Furthermore, as $(\xi, x) \mapsto \xi' V(x)\xi$ is bounded away from zero on $\mathbb{S} \times S_*$, Lemma C.2 implies that

$$\sup_{x_i \in S_*} |\hat{V}^{-1}(x_i) - V^{-1}(x_i)| \overset{\text{elt. by elt.}}{=} O_p(\tau_n).$$

Therefore, applying Lemma C.1 twice, we get

$$\sup_{x_i \in S_*} |\hat{V}^{-1}(x_i)\hat{h}^{-2}(x_i) - V^{-1}(x_i)h^{-2}(x_i)| \overset{\text{elt. by elt.}}{=} O_p(\tau_n);$$

i.e. $\sup_{x_i \in S_*} |\hat{H}^{-1}(x_i) - H^{-1}(x_i)| \overset{\text{elt. by elt.}}{=} O_p(\tau_n)$. □

**Lemma B.11.** *Let* $\frac{n^{1-2/m}b_n^s}{\log n} \to \infty$ *and suppose that the assumptions in Newey (1994, Lemma B.1, Lemma B.2) are satisfied. Then*

$$\sup_{x_i \in S_*} |\hat{H}^{-1}(x_i) - \frac{\{\mathbb{E}\hat{\Omega}(x_i)\}^{-1}}{\mathbb{E}\hat{h}(x_i)}| \overset{\text{elt. by elt.}}{=} O_p(\sqrt{\frac{\log n}{nb_n^s}}).$$

*Proof.* Let $c_n \overset{\text{def}}{=} \sqrt{\frac{\log n}{nb_n^s}}$. Newey (1994, Lemmas B.1 and B.2) shows that

$$\sup_{x_i \in S_*} |\hat{\Omega}(x_i) - \mathbb{E}\hat{\Omega}(x_i)| \overset{\text{elt. by elt.}}{=} O_p(c_n),$$

$$\sup_{x_i \in S_*} |\hat{h}(x_i) - \mathbb{E}\hat{h}(x_i)| = O_p(c_n),$$

$$\sup_{x_i \in S_*} |\mathbb{E}\hat{h}(x_i) - h(x_i)| = O_p(b_n^2).$$

Because $h$ is bounded away from zero on $S_*$, the last result implies that $\mathbb{E}\hat{h}(x_i)$ is also bounded away from zero on $S_*$ for large enough $n$. Therefore, using Lemma C.1 and the first two results, we get

$$\sup_{x_i \in S_*} |\hat{V}(x_i) - \frac{\mathbb{E}\hat{\Omega}(x_i)}{\mathbb{E}\hat{h}(x_i)}| \overset{\text{elt. by elt.}}{=} O_p(c_n).$$

In a similar manner we can show that for large enough $n$ the map $(\xi, x_i) \mapsto \xi' \mathbb{E}\hat{\Omega}(x_i)\xi$ is bounded away from zero on $\mathbb{S} \times S_*$. Therefore, we can use Lemma C.2 to show that

$$\sup_{x_i \in S_*} |\hat{V}^{-1}(x_i) - \{\frac{\mathbb{E}\hat{\Omega}(x_i)}{\mathbb{E}\hat{h}(x_i)}\}^{-1}| \overset{\text{elt. by elt.}}{=} O_p(c_n).$$

Finally, two successive applications of Lemma C.1 lead to

$$\sup_{x_i \in S_*} |\frac{\hat{V}^{-1}(x_i)}{\hat{h}^2(x_i)} - \{\frac{\mathbb{E}\hat{\Omega}(x_i)}{\mathbb{E}\hat{h}(x_i)}\}^{-1} \frac{1}{\mathbb{E}^2\hat{h}(x_i)}| \overset{\text{elt. by elt.}}{=} O_p(c_n);$$

i.e. $\sup_{x_i \in S_*} |\hat{H}^{-1}(x_i) - \frac{\{\mathbb{E}\hat{\Omega}(x_i)\}^{-1}}{\mathbb{E}\hat{h}(x_i)}| \overset{\text{elt. by elt.}}{=} O_p(c_n)$. ☐

## Appendix C. Some Useful Results

**Lemma C.1.** *Let $a_n$ and $b_n$ be sequences of positive numbers such that $a_n, b_n \downarrow 0$. Also let $r_n$ be a sequence of functions such that $\sup_x |r_n(x) - r(x)| = O_p(a_n)$ and $\sup_x |r(x)| < \infty$. Furthermore, $s_n$ is a sequence of functions such that $\sup_x |s_n(x) - s(x)| = O_p(b_n)$ and $\inf_x |s(x)| > 0$. Then $\sup_x |\frac{r_n(x)}{s_n(x)} - \frac{r(x)}{s(x)}| = O_p(\max\{a_n, b_n\})$.*

**Proof of Lemma C.1**. Let $\inf_x |s(x)| = m$, $\sup_x |r(x)| = M$, and notice that $|\inf_x |s_n(x)| - \inf_x |s(x)|| \leq \sup_x |s_n(x) - s(x)|$. Therefore, since $\sup_x |s_n(x) - s(x)| = o_p(1)$ by assumption, $\inf_x |s_n(x)| > \frac{m}{2}$ holds with probability approaching one as $n \uparrow \infty$. Hence we can use the identity $\frac{a}{b} = \frac{a}{c} - \frac{a(b-c)}{c^2} - \frac{a(b-c)^2}{bc^2}$ to show that

$$\begin{aligned}
|\frac{r_n(x)}{s_n(x)} - \frac{r(x)}{s(x)}| &\leq \frac{\sup_x |r_n(x) - r(x)|}{|s(x)|} + |r_n(x)| \frac{\sup_x |s_n(x) - s(x)|}{|s(x)|^2} \\
&\quad + \frac{|r_n(x)|}{|s_n(x)|} \frac{\{\sup_x |s_n(x) - s(x)|\}^2}{|s(x)|^2} \\
&\leq \frac{\sup_x |r_n(x) - r(x)|}{m} + |r_n(x)| \frac{\sup_x |s_n(x) - s(x)|}{m^2} \\
&\quad + 2|r_n(x)| \frac{\{\sup_x |s_n(x) - s(x)|\}^2}{m^3}
\end{aligned}$$

holds with probability approaching one. Because $|r_n(x)| \leq M + \sup_x |r_n(x) - r(x)|$, and $\sup_x |s_n(x) - s(x)| = o_p(1)$ by assumption, we get that

$$|\frac{r_n(x)}{s_n(x)} - \frac{r(x)}{s(x)}| \leq \frac{\sup_x |r_n(x) - r(x)|}{m} + M(1 + \frac{2}{m})\frac{\sup_x |s_n(x) - s(x)|}{m^2} + (1 + \frac{2}{m})\frac{\sup_x |r_n(x) - r(x)| \sup_x |s_n(x) - s(x)|}{m^2}.$$

holds with probability approaching one. The desired result follows. □

**Lemma C.2.** *Let $a_n$ be a sequence of positive numbers such that $a_n \to 0$. Also, let $\hat{V}(x)$ be a sequence of $d \times d$ symmetric positive semidefinite matrices such that $\sup_{x \in S_*} \|\hat{V}(x) - V(x)\| = O_p(a_n)$, and $(\xi, x) \mapsto \xi' V(x)\xi$ is bounded away from zero on $\mathbb{S} \times S_*$. Then $\sup_{x \in S_*} \|\hat{V}^{-1}(x) - V^{-1}(x)\| = O_p(a_n)$.*

**Proof of Lemma C.2**. Pick any $\alpha \in \mathbb{S}$. Clearly,

$$\sup_{(\alpha,x) \in \mathbb{S} \times S_*} |\alpha' \hat{V}(x)\alpha - \alpha' V(x)\alpha| = O_p(a_n).$$

Since $\alpha' V(x)\alpha$ is bounded away from zero on $\mathbb{S} \times S_*$, we can use Lemma C.1 to see that $\sup_{(\alpha,x) \in \mathbb{S} \times S_*} |\frac{1}{\alpha' \hat{V}(x)\alpha} - \frac{1}{\alpha' V(x)\alpha}| = O_p(a_n)$. This means that for any $\xi \in \mathbb{S}$, we can write

$$\sup_{(\alpha,x) \in \mathbb{S} \times S_*} |\frac{(\alpha'\xi)^2}{\alpha' \hat{V}(x)\alpha} - \frac{(\alpha'\xi)^2}{\alpha' V(x)\alpha}| = O_p(a_n).$$

In particular, this implies that

$$\sup_{x \in S_*} |\sup_{\alpha \in \mathbb{S}} \frac{(\alpha'\xi)^2}{\alpha' \hat{V}(x)\alpha} - \sup_{\alpha \in \mathbb{S}} \frac{(\alpha'\xi)^2}{\alpha' V(x)\alpha}| = O_p(a_n),$$

which, for large enough $n$, reduces to

$$\sup_{x \in S_*} |\xi' \hat{V}^{-1}(x)\xi - \xi' V^{-1}(x)\xi| = O_p(a_n).$$

As $\xi \in \mathbb{S}$ was arbitrary we obtain the required result. □

**Lemma C.3.** *Let $z_1, \ldots, z_n$ be a sequence of iid random vectors such that $\mathbb{E}\|z_1\|^m < \infty$ for some $m > 0$, and let $z_n^* = \max\{\|z_1\|, \ldots, \|z_n\|\}$. Then $\Pr\{z_n^* < n^{1/m}\} = 1$ for large enough $n$.*

**Proof of Lemma C.3**. For the reader's convenience we repeat the argument in Owen (1988, Page 241). Since $\sum_{n=1}^{\infty} \Pr\{\|z_1\|^m \geq n\} \leq \mathbb{E}\|z_1\|^m$, we have $\sum_{n=1}^{\infty} \Pr\{\|z_1\|^m \geq n\} < \infty$. But because $z_1, \ldots, z_n$ are identically distributed, $\sum_{n=1}^{\infty} \Pr\{\|z_n\|^m \geq n\} < \infty$. Therefore, by the Borel-Cantelli lemma $\Pr\{A_n \ i.o.\} = 0$, where $A_n \stackrel{\text{def}}{=} \{\|z_n\|^m \geq n\}$. Thus the event $\{\|z_n\| < n^{1/m}\}$ happens for all but finitely many $n$ with probability one. Since $n^{1/m}$ will eventually exceed the largest element in the finite collection of $\|z_k\|$'s that exceed $k^{1/m}$, we get $\Pr\{z_n^* < n^{1/m}\} = 1$ when $n$ is large enough. □

## Appendix D. Some Useful Summation Identities

The following identities have been used throughout the paper.

$$\text{(D.1)}\quad \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\sum_{k=1}^{n} b_{ik} = \sum_{i=1}^{n} a_{ii}b_{ii} + \sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} a_{ij}b_{ij} + \sum_{i=1}^{n}\sum_{t=1,t\neq i}^{n} a_{ii}b_{it} + \sum_{i=1}^{n}\sum_{k=1,k\neq i}^{n} a_{ik}b_{ii} + \sum_{i=1}^{n}\sum_{k=1,k\neq i}^{n}\sum_{t=1,t\neq k\neq i}^{n} a_{ik}b_{it}$$

$$\text{(D.2)}\quad \{\sum_{t=1}^{n}\sum_{j=1,j\neq t}^{n} a_{tj}\}^2 = \sum_{t=1}^{n}\sum_{j=1,j\neq t}^{n} a_{tj}^2 + \sum_{t=1}^{n}\sum_{j=1,j\neq t}^{n}\sum_{k=1,k\neq j\neq t}^{n} a_{tj}a_{tk} + \sum_{t=1}^{n}\sum_{j=1,j\neq t}^{n}\sum_{u=1,u\neq j\neq t}^{n} a_{tj}a_{ju} + \sum_{t=1}^{n}\sum_{j=1,j\neq t}^{n} a_{tj}a_{jt} + \sum_{t=1}^{n}\sum_{j=1,j\neq t}^{n}\sum_{m=1,m\neq j\neq t}^{n} a_{tj}a_{mj} + \sum_{t=1}^{n}\sum_{j=1,j\neq t}^{n}\sum_{m=1,m\neq j\neq t}^{n} a_{tj}a_{mt} + \sum_{t=1}^{n}\sum_{j=1,j\neq t}^{n}\sum_{m=1,m\neq j\neq t}^{n}\sum_{u=1,u\neq m\neq j\neq t}^{n} a_{tj}a_{mu}$$

$$\text{(D.3)}\quad \{\sum_{t=1}^{n-1}\sum_{j=t+1}^{n} a_{tj}\}^2 = \sum_{t=1}^{n-1}\sum_{j=t+1}^{n} a_{tj}^2 + \sum_{t=1}^{n-2}\sum_{j=t+1}^{n}\sum_{k=t+1,k\neq j}^{n} a_{tj}a_{tk} + \sum_{t=1}^{n-1}\sum_{j=t+1}^{n}\sum_{m=1,m\neq t}^{n-1}\sum_{u=m+1}^{n} a_{tj}a_{mu}$$

$$\text{(D.4)}\quad \{\sum_{i=1}^{n} a_i\}^4 = \sum_{i=1}^{n} a_i^4 + 4\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} a_i a_j^3 + 3\sum_{i=1}^{n}\sum_{k=1,k\neq i}^{n} a_i^2 a_k^2 + 6\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\sum_{m=1,m\neq j\neq i}^{n} a_i a_j^2 a_m + \sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\sum_{l=1,l\neq j\neq i}^{n}\sum_{m=1,m\neq l\neq j\neq i}^{n} a_i a_j a_l a_m$$

$\square$

## Appendix E. Simulation Results

Table 1. $z = c(1 + x^2)^{1/2} + \varepsilon(1 + x^2)^{1/2}$

| | | Size Corrected | | | Size Uncorrected | | |
|---|---|---|---|---|---|---|---|
| $n = 100$ | | 1% | 5% | 10% | 1% | 5% | 10% |
| $c = 0.0$ | SELR | 0.007 | 0.049 | 0.103 | 0.019 | 0.063 | 0.103 |
| | $\text{ABS}_{\text{uw}}$ | 0.006 | 0.056 | 0.098 | 0.010 | 0.048 | 0.073 |
| | Zheng | 0.008 | 0.049 | 0.091 | 0.024 | 0.066 | 0.104 |
| $c = 0.1$ | SELR | 0.026 | 0.103 | 0.174 | 0.052 | 0.120 | 0.176 |
| | $\text{ABS}_{\text{uw}}$ | 0.030 | 0.093 | 0.162 | 0.038 | 0.089 | 0.126 |
| | Zheng | 0.024 | 0.082 | 0.154 | 0.045 | 0.109 | 0.164 |
| $c = 0.2$ | SELR | 0.141 | 0.311 | 0.423 | 0.207 | 0.348 | 0.428 |
| | $\text{ABS}_{\text{uw}}$ | 0.122 | 0.290 | 0.399 | 0.144 | 0.263 | 0.359 |
| | Zheng | 0.099 | 0.242 | 0.369 | 0.171 | 0.294 | 0.383 |
| $c = 0.3$ | SELR | 0.430 | 0.639 | 0.728 | 0.520 | 0.664 | 0.729 |
| | $\text{ABS}_{\text{uw}}$ | 0.387 | 0.614 | 0.716 | 0.422 | 0.577 | 0.672 |
| | Zheng | 0.318 | 0.552 | 0.682 | 0.450 | 0.598 | 0.694 |
| $c = 0.4$ | SELR | 0.738 | 0.879 | 0.936 | 0.815 | 0.895 | 0.937 |
| | $\text{ABS}_{\text{uw}}$ | 0.707 | 0.864 | 0.927 | 0.733 | 0.851 | 0.901 |
| | Zheng | 0.639 | 0.825 | 0.895 | 0.753 | 0.861 | 0.901 |
| $c = 0.5$ | SELR | 0.941 | 0.979 | 0.990 | 0.963 | 0.982 | 0.990 |
| | $\text{ABS}_{\text{uw}}$ | 0.928 | 0.974 | 0.986 | 0.937 | 0.971 | 0.982 |
| | Zheng | 0.881 | 0.963 | 0.979 | 0.942 | 0.970 | 0.981 |
| | | Size Corrected | | | Size Uncorrected | | |
| $n = 250$ | | 1% | 5% | 10% | 1% | 5% | 10% |
| $c = 0.0$ | SELR | 0.013 | 0.055 | 0.097 | 0.026 | 0.068 | 0.101 |
| | $\text{ABS}_{\text{uw}}$ | 0.009 | 0.049 | 0.093 | 0.026 | 0.053 | 0.085 |
| | Zheng | 0.010 | 0.052 | 0.096 | 0.028 | 0.072 | 0.102 |
| $c = 0.1$ | SELR | 0.063 | 0.186 | 0.298 | 0.126 | 0.234 | 0.301 |
| | $\text{ABS}_{\text{uw}}$ | 0.062 | 0.190 | 0.297 | 0.094 | 0.199 | 0.273 |
| | Zheng | 0.050 | 0.161 | 0.265 | 0.096 | 0.208 | 0.278 |
| $c = 0.2$ | SELR | 0.433 | 0.645 | 0.756 | 0.567 | 0.689 | 0.762 |
| | $\text{ABS}_{\text{uw}}$ | 0.407 | 0.640 | 0.753 | 0.504 | 0.655 | 0.726 |
| | Zheng | 0.346 | 0.592 | 0.701 | 0.494 | 0.637 | 0.715 |
| $c = 0.3$ | SELR | 0.889 | 0.970 | 0.989 | 0.943 | 0.980 | 0.990 |
| | $\text{ABS}_{\text{uw}}$ | 0.871 | 0.966 | 0.985 | 0.928 | 0.969 | 0.982 |
| | Zheng | 0.831 | 0.947 | 0.969 | 0.910 | 0.952 | 0.971 |
| $c = 0.4$ | SELR | 0.994 | 0.997 | 0.999 | 0.997 | 0.998 | 0.999 |
| | $\text{ABS}_{\text{uw}}$ | 0.994 | 0.998 | 0.998 | 0.996 | 0.998 | 0.998 |
| | Zheng | 0.989 | 0.998 | 1.000 | 0.997 | 0.998 | 1.000 |
| $c = 0.5$ | SELR | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | $\text{ABS}_{\text{uw}}$ | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | Zheng | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

Table 2. $z = c(1 + x^2) + \varepsilon(1 + x^2)^{1/2}$

| | | Size Corrected | | | Size Uncorrected | | |
|---|---|---|---|---|---|---|---|
| $n = 100$ | | 1% | 5% | 10% | 1% | 5% | 10% |
| $c = 0.0$ | SELR | 0.007 | 0.049 | 0.103 | 0.019 | 0.063 | 0.103 |
| | $\text{ABS}_{\text{uw}}$ | 0.006 | 0.056 | 0.098 | 0.010 | 0.048 | 0.073 |
| | Zheng | 0.008 | 0.049 | 0.091 | 0.024 | 0.066 | 0.104 |
| $c = 0.1$ | SELR | 0.034 | 0.126 | 0.204 | 0.069 | 0.148 | 0.206 |
| | $\text{ABS}_{\text{uw}}$ | 0.041 | 0.106 | 0.196 | 0.048 | 0.098 | 0.157 |
| | Zheng | 0.028 | 0.103 | 0.176 | 0.058 | 0.134 | 0.188 |
| $c = 0.2$ | SELR | 0.210 | 0.405 | 0.527 | 0.297 | 0.443 | 0.529 |
| | $\text{ABS}_{\text{uw}}$ | 0.195 | 0.395 | 0.513 | 0.223 | 0.365 | 0.462 |
| | Zheng | 0.151 | 0.344 | 0.475 | 0.257 | 0.398 | 0.497 |
| $c = 0.3$ | SELR | 0.579 | 0.752 | 0.838 | 0.674 | 0.778 | 0.839 |
| | $\text{ABS}_{\text{uw}}$ | 0.545 | 0.751 | 0.832 | 0.588 | 0.731 | 0.792 |
| | Zheng | 0.484 | 0.700 | 0.809 | 0.606 | 0.749 | 0.820 |
| $c = 0.4$ | SELR | 0.885 | 0.959 | 0.975 | 0.922 | 0.962 | 0.975 |
| | $\text{ABS}_{\text{uw}}$ | 0.866 | 0.949 | 0.976 | 0.885 | 0.945 | 0.964 |
| | Zheng | 0.831 | 0.932 | 0.961 | 0.892 | 0.952 | 0.964 |
| $c = 0.5$ | SELR | 0.986 | 0.999 | 0.999 | 0.993 | 0.999 | 0.999 |
| | $\text{ABS}_{\text{uw}}$ | 0.983 | 0.998 | 0.999 | 0.986 | 0.998 | 0.999 |
| | Zheng | 0.973 | 0.993 | 0.999 | 0.988 | 0.995 | 0.999 |
| | | Size Corrected | | | Size Uncorrected | | |
| $n = 250$ | | 1% | 5% | 10% | 1% | 5% | 10% |
| $c = 0.0$ | SELR | 0.013 | 0.055 | 0.097 | 0.026 | 0.068 | 0.101 |
| | $\text{ABS}_{\text{uw}}$ | 0.009 | 0.049 | 0.093 | 0.026 | 0.053 | 0.085 |
| | Zheng | 0.010 | 0.052 | 0.096 | 0.028 | 0.072 | 0.102 |
| $c = 0.1$ | SELR | 0.096 | 0.244 | 0.354 | 0.169 | 0.284 | 0.366 |
| | $\text{ABS}_{\text{uw}}$ | 0.084 | 0.245 | 0.362 | 0.145 | 0.257 | 0.339 |
| | Zheng | 0.069 | 0.224 | 0.327 | 0.146 | 0.264 | 0.346 |
| $c = 0.2$ | SELR | 0.599 | 0.788 | 0.872 | 0.709 | 0.819 | 0.879 |
| | $\text{ABS}_{\text{uw}}$ | 0.579 | 0.804 | 0.882 | 0.677 | 0.809 | 0.866 |
| | Zheng | 0.534 | 0.754 | 0.834 | 0.663 | 0.790 | 0.850 |
| $c = 0.3$ | SELR | 0.975 | 0.994 | 0.995 | 0.989 | 0.995 | 0.995 |
| | $\text{ABS}_{\text{uw}}$ | 0.970 | 0.994 | 0.996 | 0.980 | 0.994 | 0.996 |
| | Zheng | 0.953 | 0.987 | 0.994 | 0.977 | 0.989 | 0.994 |
| $c = 0.4$ | SELR | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | $\text{ABS}_{\text{uw}}$ | 0.999 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 |
| | Zheng | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $c = 0.5$ | SELR | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | $\text{ABS}_{\text{uw}}$ | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | Zheng | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

TABLE 3. $z = cx^{1/2} + \varepsilon x^{1/2}$

| $n = 100$ | | Size Corrected 1% | 5% | 10% | Size Uncorrected 1% | 5% | 10% |
|---|---|---|---|---|---|---|---|
| $c = 0.0$ | SELR | 0.008 | 0.054 | 0.104 | 0.013 | 0.047 | 0.078 |
| | $\mathrm{ABS_{uw}}$ | 0.007 | 0.053 | 0.095 | 0.016 | 0.050 | 0.077 |
| | Zheng | 0.006 | 0.047 | 0.091 | 0.030 | 0.068 | 0.100 |
| $c = 0.1$ | SELR | 0.026 | 0.099 | 0.174 | 0.037 | 0.089 | 0.140 |
| | $\mathrm{ABS_{uw}}$ | 0.025 | 0.081 | 0.159 | 0.037 | 0.077 | 0.124 |
| | Zheng | 0.020 | 0.073 | 0.140 | 0.046 | 0.098 | 0.153 |
| $c = 0.2$ | SELR | 0.129 | 0.298 | 0.409 | 0.158 | 0.282 | 0.363 |
| | $\mathrm{ABS_{uw}}$ | 0.090 | 0.249 | 0.375 | 0.132 | 0.236 | 0.316 |
| | Zheng | 0.078 | 0.202 | 0.325 | 0.152 | 0.249 | 0.344 |
| $c = 0.3$ | SELR | 0.392 | 0.610 | 0.706 | 0.438 | 0.588 | 0.671 |
| | $\mathrm{ABS_{uw}}$ | 0.288 | 0.529 | 0.669 | 0.367 | 0.516 | 0.616 |
| | Zheng | 0.245 | 0.470 | 0.609 | 0.389 | 0.542 | 0.626 |
| $c = 0.4$ | SELR | 0.710 | 0.867 | 0.926 | 0.752 | 0.856 | 0.903 |
| | $\mathrm{ABS_{uw}}$ | 0.598 | 0.802 | 0.877 | 0.670 | 0.791 | 0.851 |
| | Zheng | 0.536 | 0.751 | 0.846 | 0.686 | 0.813 | 0.857 |
| $c = 0.5$ | SELR | 0.921 | 0.974 | 0.985 | 0.940 | 0.970 | 0.980 |
| | $\mathrm{ABS_{uw}}$ | 0.850 | 0.949 | 0.976 | 0.880 | 0.945 | 0.971 |
| | Zheng | 0.802 | 0.918 | 0.964 | 0.895 | 0.944 | 0.966 |

| $n = 250$ | | Size Corrected 1% | 5% | 10% | Size Uncorrected 1% | 5% | 10% |
|---|---|---|---|---|---|---|---|
| $c = 0.0$ | SELR | 0.013 | 0.054 | 0.092 | 0.021 | 0.053 | 0.074 |
| | $\mathrm{ABS_{uw}}$ | 0.011 | 0.049 | 0.093 | 0.027 | 0.052 | 0.084 |
| | Zheng | 0.010 | 0.044 | 0.096 | 0.028 | 0.062 | 0.104 |
| $c = 0.1$ | SELR | 0.059 | 0.191 | 0.287 | 0.096 | 0.188 | 0.263 |
| | $\mathrm{ABS_{uw}}$ | 0.052 | 0.169 | 0.273 | 0.081 | 0.175 | 0.243 |
| | Zheng | 0.043 | 0.137 | 0.237 | 0.090 | 0.179 | 0.250 |
| $c = 0.2$ | SELR | 0.415 | 0.641 | 0.746 | 0.500 | 0.639 | 0.716 |
| | $\mathrm{ABS_{uw}}$ | 0.334 | 0.590 | 0.701 | 0.440 | 0.598 | 0.683 |
| | Zheng | 0.264 | 0.509 | 0.648 | 0.421 | 0.575 | 0.658 |
| $c = 0.3$ | SELR | 0.877 | 0.969 | 0.987 | 0.917 | 0.969 | 0.983 |
| | $\mathrm{ABS_{uw}}$ | 0.810 | 0.944 | 0.972 | 0.891 | 0.951 | 0.967 |
| | Zheng | 0.739 | 0.909 | 0.950 | 0.864 | 0.929 | 0.958 |
| $c = 0.4$ | SELR | 0.994 | 0.998 | 0.999 | 0.996 | 0.998 | 0.998 |
| | $\mathrm{ABS_{uw}}$ | 0.989 | 0.996 | 0.998 | 0.992 | 0.996 | 0.997 |
| | Zheng | 0.978 | 0.993 | 0.998 | 0.989 | 0.996 | 0.998 |
| $c = 0.5$ | SELR | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | $\mathrm{ABS_{uw}}$ | 0.999 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 |
| | Zheng | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

Table 4. $z = cx + \varepsilon x^{1/2}$

| | | Size Corrected | | | Size Uncorrected | | |
|---|---|---|---|---|---|---|---|
| $n = 100$ | | 1% | 5% | 10% | 1% | 5% | 10% |
| $c = 0.0$ | SELR | 0.008 | 0.054 | 0.104 | 0.013 | 0.047 | 0.078 |
| | $\text{ABS}_{\text{uw}}$ | 0.007 | 0.053 | 0.095 | 0.016 | 0.050 | 0.077 |
| | Zheng | 0.006 | 0.047 | 0.091 | 0.030 | 0.068 | 0.100 |
| $c = 0.1$ | SELR | 0.013 | 0.074 | 0.144 | 0.024 | 0.069 | 0.107 |
| | $\text{ABS}_{\text{uw}}$ | 0.021 | 0.071 | 0.134 | 0.031 | 0.065 | 0.094 |
| | Zheng | 0.016 | 0.059 | 0.126 | 0.041 | 0.080 | 0.137 |
| $c = 0.2$ | SELR | 0.067 | 0.176 | 0.275 | 0.083 | 0.167 | 0.234 |
| | $\text{ABS}_{\text{uw}}$ | 0.063 | 0.170 | 0.270 | 0.075 | 0.161 | 0.225 |
| | Zheng | 0.046 | 0.141 | 0.233 | 0.095 | 0.187 | 0.244 |
| $c = 0.3$ | SELR | 0.182 | 0.361 | 0.475 | 0.225 | 0.343 | 0.429 |
| | $\text{ABS}_{\text{uw}}$ | 0.155 | 0.346 | 0.472 | 0.202 | 0.327 | 0.423 |
| | Zheng | 0.130 | 0.304 | 0.435 | 0.235 | 0.360 | 0.452 |
| $c = 0.4$ | SELR | 0.368 | 0.599 | 0.691 | 0.432 | 0.574 | 0.649 |
| | $\text{ABS}_{\text{uw}}$ | 0.348 | 0.581 | 0.703 | 0.414 | 0.560 | 0.650 |
| | Zheng | 0.303 | 0.529 | 0.659 | 0.444 | 0.589 | 0.673 |
| $c = 0.5$ | SELR | 0.609 | 0.797 | 0.868 | 0.658 | 0.785 | 0.837 |
| | $\text{ABS}_{\text{uw}}$ | 0.573 | 0.788 | 0.865 | 0.638 | 0.779 | 0.828 |
| | Zheng | 0.530 | 0.750 | 0.833 | 0.677 | 0.797 | 0.839 |
| | | Size Corrected | | | Size Uncorrected | | |
| $n = 250$ | | 1% | 5% | 10% | 1% | 5% | 10% |
| $c = 0.0$ | SELR | 0.013 | 0.054 | 0.092 | 0.021 | 0.053 | 0.074 |
| | $\text{ABS}_{\text{uw}}$ | 0.011 | 0.049 | 0.093 | 0.027 | 0.052 | 0.082 |
| | Zheng | 0.010 | 0.044 | 0.096 | 0.028 | 0.062 | 0.104 |
| $c = 0.1$ | SELR | 0.037 | 0.115 | 0.201 | 0.054 | 0.115 | 0.173 |
| | $\text{ABS}_{\text{uw}}$ | 0.039 | 0.123 | 0.197 | 0.062 | 0.128 | 0.176 |
| | Zheng | 0.030 | 0.106 | 0.189 | 0.071 | 0.137 | 0.201 |
| $c = 0.2$ | SELR | 0.182 | 0.384 | 0.506 | 0.232 | 0.379 | 0.480 |
| | $\text{ABS}_{\text{uw}}$ | 0.187 | 0.394 | 0.505 | 0.261 | 0.400 | 0.480 |
| | Zheng | 0.156 | 0.346 | 0.459 | 0.255 | 0.406 | 0.478 |
| $c = 0.3$ | SELR | 0.519 | 0.730 | 0.825 | 0.596 | 0.727 | 0.803 |
| | $\text{ABS}_{\text{uw}}$ | 0.517 | 0.742 | 0.851 | 0.620 | 0.758 | 0.836 |
| | Zheng | 0.460 | 0.693 | 0.810 | 0.609 | 0.749 | 0.816 |
| $c = 0.4$ | SELR | 0.856 | 0.958 | 0.973 | 0.907 | 0.958 | 0.971 |
| | $\text{ABS}_{\text{uw}}$ | 0.863 | 0.960 | 0.980 | 0.921 | 0.961 | 0.976 |
| | Zheng | 0.820 | 0.939 | 0.971 | 0.908 | 0.956 | 0.972 |
| $c = 0.5$ | SELR | 0.982 | 0.995 | 0.996 | 0.990 | 0.995 | 0.996 |
| | $\text{ABS}_{\text{uw}}$ | 0.982 | 0.994 | 0.997 | 0.991 | 0.994 | 0.995 |
| | Zheng | 0.971 | 0.993 | 0.997 | 0.990 | 0.994 | 0.998 |

Table 5. $z = c(0.5 + I\{x \geq 0.5\})^{1/2} + \varepsilon(0.5 + I\{x \geq 0.5\})^{1/2}$

| | | Size Corrected | | | Size Uncorrected | | |
|---|---|---|---|---|---|---|---|
| $n = 100$ | | 1% | 5% | 10% | 1% | 5% | 10% |
| $c = 0.0$ | SELR | 0.011 | 0.047 | 0.100 | 0.016 | 0.044 | 0.088 |
| | $\text{ABS}_{\text{uw}}$ | 0.009 | 0.050 | 0.097 | 0.014 | 0.047 | 0.074 |
| | Zheng | 0.008 | 0.047 | 0.094 | 0.030 | 0.062 | 0.100 |
| $c = 0.1$ | SELR | 0.026 | 0.103 | 0.174 | 0.044 | 0.101 | 0.147 |
| | $\text{ABS}_{\text{uw}}$ | 0.026 | 0.087 | 0.160 | 0.033 | 0.084 | 0.127 |
| | Zheng | 0.021 | 0.069 | 0.145 | 0.044 | 0.104 | 0.161 |
| $c = 0.2$ | SELR | 0.125 | 0.305 | 0.415 | 0.171 | 0.303 | 0.372 |
| | $\text{ABS}_{\text{uw}}$ | 0.093 | 0.248 | 0.383 | 0.131 | 0.238 | 0.328 |
| | Zheng | 0.082 | 0.212 | 0.334 | 0.162 | 0.267 | 0.351 |
| $c = 0.3$ | SELR | 0.385 | 0.617 | 0.724 | 0.463 | 0.613 | 0.699 |
| | $\text{ABS}_{\text{uw}}$ | 0.309 | 0.540 | 0.668 | 0.369 | 0.528 | 0.622 |
| | Zheng | 0.257 | 0.500 | 0.632 | 0.413 | 0.566 | 0.645 |
| $c = 0.4$ | SELR | 0.709 | 0.872 | 0.929 | 0.782 | 0.871 | 0.914 |
| | $\text{ABS}_{\text{uw}}$ | 0.613 | 0.810 | 0.893 | 0.679 | 0.798 | 0.867 |
| | Zheng | 0.574 | 0.773 | 0.864 | 0.700 | 0.824 | 0.868 |
| $c = 0.5$ | SELR | 0.918 | 0.976 | 0.993 | 0.947 | 0.976 | 0.988 |
| | $\text{ABS}_{\text{uw}}$ | 0.865 | 0.954 | 0.984 | 0.890 | 0.950 | 0.974 |
| | Zheng | 0.824 | 0.932 | 0.968 | 0.903 | 0.954 | 0.968 |

| | | Size Corrected | | | Size Uncorrected | | |
|---|---|---|---|---|---|---|---|
| $n = 250$ | | 1% | 5% | 10% | 1% | 5% | 10% |
| $c = 0.0$ | SELR | 0.014 | 0.054 | 0.095 | 0.020 | 0.058 | 0.082 |
| | $\text{ABS}_{\text{uw}}$ | 0.010 | 0.049 | 0.098 | 0.023 | 0.053 | 0.088 |
| | Zheng | 0.009 | 0.047 | 0.093 | 0.021 | 0.068 | 0.105 |
| $c = 0.1$ | SELR | 0.060 | 0.187 | 0.292 | 0.095 | 0.199 | 0.270 |
| | $\text{ABS}_{\text{uw}}$ | 0.051 | 0.164 | 0.257 | 0.087 | 0.171 | 0.235 |
| | Zheng | 0.046 | 0.143 | 0.229 | 0.093 | 0.180 | 0.239 |
| $c = 0.2$ | SELR | 0.420 | 0.641 | 0.747 | 0.515 | 0.655 | 0.732 |
| | $\text{ABS}_{\text{uw}}$ | 0.327 | 0.586 | 0.697 | 0.455 | 0.592 | 0.676 |
| | Zheng | 0.289 | 0.538 | 0.652 | 0.452 | 0.594 | 0.667 |
| $c = 0.3$ | SELR | 0.884 | 0.969 | 0.985 | 0.925 | 0.971 | 0.982 |
| | $\text{ABS}_{\text{uw}}$ | 0.815 | 0.948 | 0.971 | 0.891 | 0.949 | 0.969 |
| | Zheng | 0.766 | 0.916 | 0.953 | 0.873 | 0.936 | 0.958 |
| $c = 0.4$ | SELR | 0.995 | 0.998 | 0.999 | 0.996 | 0.998 | 0.999 |
| | $\text{ABS}_{\text{uw}}$ | 0.985 | 0.996 | 0.998 | 0.991 | 0.996 | 0.998 |
| | Zheng | 0.980 | 0.992 | 0.998 | 0.989 | 0.996 | 0.998 |
| $c = 0.5$ | SELR | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | $\text{ABS}_{\text{uw}}$ | 0.999 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 |
| | Zheng | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

TABLE 6. $z = c(0.5 + I\{x \geq 0.5\}) + \varepsilon(0.5 + I\{x \geq 0.5\})^{1/2}$

| | | Size Corrected | | | Size Uncorrected | | |
|---|---|---|---|---|---|---|---|
| $n = 100$ | | 1% | 5% | 10% | 1% | 5% | 10% |
| $c = 0.0$ | SELR | 0.011 | 0.047 | 0.100 | 0.016 | 0.044 | 0.088 |
| | $\text{ABS}_{\text{uw}}$ | 0.009 | 0.050 | 0.097 | 0.014 | 0.047 | 0.074 |
| | Zheng | 0.008 | 0.047 | 0.094 | 0.030 | 0.062 | 0.100 |
| $c = 0.1$ | SELR | 0.028 | 0.115 | 0.181 | 0.048 | 0.114 | 0.157 |
| | $\text{ABS}_{\text{uw}}$ | 0.031 | 0.092 | 0.180 | 0.044 | 0.088 | 0.135 |
| | Zheng | 0.025 | 0.082 | 0.158 | 0.054 | 0.121 | 0.166 |
| $c = 0.2$ | SELR | 0.142 | 0.329 | 0.429 | 0.193 | 0.327 | 0.397 |
| | $\text{ABS}_{\text{uw}}$ | 0.129 | 0.304 | 0.445 | 0.173 | 0.297 | 0.385 |
| | Zheng | 0.111 | 0.255 | 0.402 | 0.198 | 0.322 | 0.425 |
| $c = 0.3$ | SELR | 0.421 | 0.640 | 0.737 | 0.492 | 0.637 | 0.708 |
| | $\text{ABS}_{\text{uw}}$ | 0.394 | 0.622 | 0.726 | 0.467 | 0.611 | 0.695 |
| | Zheng | 0.358 | 0.576 | 0.710 | 0.494 | 0.638 | 0.720 |
| $c = 0.4$ | SELR | 0.735 | 0.893 | 0.930 | 0.805 | 0.890 | 0.918 |
| | $\text{ABS}_{\text{uw}}$ | 0.704 | 0.874 | 0.922 | 0.753 | 0.866 | 0.904 |
| | Zheng | 0.676 | 0.847 | 0.898 | 0.787 | 0.874 | 0.905 |
| $c = 0.5$ | SELR | 0.932 | 0.982 | 0.995 | 0.958 | 0.982 | 0.992 |
| | $\text{ABS}_{\text{uw}}$ | 0.913 | 0.978 | 0.994 | 0.931 | 0.978 | 0.987 |
| | Zheng | 0.888 | 0.965 | 0.984 | 0.940 | 0.980 | 0.987 |
| | | Size Corrected | | | Size Uncorrected | | |
| $n = 250$ | | 1% | 5% | 10% | 1% | 5% | 10% |
| $c = 0.0$ | SELR | 0.014 | 0.054 | 0.095 | 0.020 | 0.058 | 0.082 |
| | $\text{ABS}_{\text{uw}}$ | 0.010 | 0.049 | 0.098 | 0.023 | 0.053 | 0.088 |
| | Zheng | 0.009 | 0.047 | 0.093 | 0.021 | 0.068 | 0.105 |
| $c = 0.1$ | SELR | 0.066 | 0.188 | 0.297 | 0.104 | 0.203 | 0.274 |
| | $\text{ABS}_{\text{uw}}$ | 0.069 | 0.192 | 0.294 | 0.120 | 0.206 | 0.280 |
| | Zheng | 0.060 | 0.173 | 0.272 | 0.114 | 0.213 | 0.289 |
| $c = 0.2$ | SELR | 0.458 | 0.660 | 0.780 | 0.533 | 0.677 | 0.754 |
| | $\text{ABS}_{\text{uw}}$ | 0.427 | 0.686 | 0.778 | 0.551 | 0.697 | 0.771 |
| | Zheng | 0.399 | 0.627 | 0.744 | 0.543 | 0.684 | 0.759 |
| $c = 0.3$ | SELR | 0.909 | 0.972 | 0.985 | 0.945 | 0.974 | 0.982 |
| | $\text{ABS}_{\text{uw}}$ | 0.895 | 0.974 | 0.983 | 0.940 | 0.975 | 0.982 |
| | Zheng | 0.872 | 0.960 | 0.978 | 0.925 | 0.969 | 0.981 |
| $c = 0.4$ | SELR | 0.995 | 0.999 | 0.999 | 0.997 | 0.999 | 0.999 |
| | $\text{ABS}_{\text{uw}}$ | 0.992 | 0.999 | 0.999 | 0.996 | 0.999 | 0.999 |
| | Zheng | 0.991 | 0.999 | 1.000 | 0.997 | 0.999 | 1.000 |
| $c = 0.5$ | SELR | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | $\text{ABS}_{\text{uw}}$ | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | Zheng | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

## Acknowledgements

We thank Don Andrews, Bruce Hansen, Wolfgang Härdle, Yongmiao Hong, Oliver Linton, Ken West, and seminar participants at the 1999 summer meetings of the econometric society, for helpful suggestions. Shane Sherlund provided excellent research assistance. The first author would also like to thank the hospitality of Professor Wolfgang Härdle at the Institute of Statistics and Econometrics, Humboldt University-Berlin, where part of this research was carried out. The second author acknowledges research support from the Alfred P. Sloan Foundation Research Fellowship and from the NSF via grants SBR-9632101 and SES-9905247.


## References


Ai, C. (1997): "A Semiparametric Maximum Likelihood Estimator," *Econometrica*, 65, 933–963.

Aït-Sahalia, Y., P. Bickel, and T. Stoker (2000): "Goodness-of-Fit Tests for Kernel Regression with an Application to Option-Implied Volatilities," Manuscript from `http://www.princeton.edu/~yacine/`.

Andrews, D. (1997): "A Conditional Kolmogorov Test," *Econometrica*, 65(5), 1097–1128.

Andrews, D. W., and W. Ploberger (1994): "Optimal tests when a nuisance parameter is present only under the alternative," *Econometrica*, 62, 1383–1414.

——— (1996): "Testing for serial correlation against an ARMA(1,1) process," *Journal of the American Statistical Association*, 435, 1331–1342.

Bierens, H. J. (1990): "A Consistent Conditional Moment Test of Functional Form," *Econometrica*, 58, 1443–1458.

——— (1994): "On the Limit Behavior of a Chi-Square Type Test if the Number of Conditional Moments Approach Infinity," *Econometric Theory*, 10, 70–90.

Bierens, H. J., and W. Ploberger (1997): "Asymptotic Theory of Integrated Conditional Moment Tests," *Econometrica*, 65(5), 1129–1151.

Brillinger, D. (1977): "Discussion of 'Consistent Nonparametric Regression' by C.J. Stone," *The Annals of Statistics*, 5, 622–623.

Brown, B. W., and W. K. Newey (1998): "Efficient Bootstrapping for Semiparametric Models," Manuscript.

de Jong, P. (1987): "A Central Limit Theorem for Generalized Quadratic Forms," *Probability Theory and Related Fields*, 75, 261–277.

Eubank, R., and C. Spiegelman (1990): "Testing the Goodness of Fit of a Linear Model via Nonparametric Regression Techniques," *Journal of the American Statistical Association*, 85, 387–392.

Fan, Y., and Q. Li (1996): "Consistent Model Specification Tests: Omitted Variables and Semiparametric Functional Forms," *Econometrica*, 64(4), 865–890.

Fisher, N. I., P. Hall, B.-Y. Jing, and A. T. Wood (1996): "Improved Pivotal Methods for Constructiong Confidence Regions with Directional Data," *Journal of the American Statistical Association*, 91, 1062–1070.

Hannan, E. (1970): *Time Series Analysis*. New York: Wiley.

Härdle, W. (1989): "Asymptotic maximal deviation of M-smoothers," *Journal of Multivariate Analysis*, 29, 163–179.

Härdle, W., and E. Mammen (1993): "Comparing Nonparametric versus Parametric Regression Fits," *Annals of Statistics*, 21, 1926–1947.

Härdle, W., and J. Marron (1990): "Semiparametric Comparison of Regression Curves," *Annals of Statistics*, 18, 63–89.

Hastie, T., and R. Tibshirani (1986): "Generalized Additive Models," *Statistical Science*, 1, 297–318.

Hong, Y., and H. White (1995): "Consistent Specification Testing via Nonparametric Series Regressions," *Econometrica*, 63, 1133–1160.

Horowitz, J. L., and V. G. Spokoiny (1999): "An Adaptive Rate-Optimal Test of a Parametric Model Against a Nonparametric Alternative," Manuscript.

Johnston, G. (1982): "Probabilities of maximal deviations for nonparametric regression function estimates," *Journal of Multivariate Analysis*, 12, 402–414.

Kitamura, Y. (1997): "Empirical likelihood methods with weakly dependent processes," *Annals of Statistics*, 25(5), 2084–2102.

——— (1999): "Asymptotic Optimality of Emptrical Likelihood for Testing Moment Restrictions," Working Paper, Department of Economics, University of Wisconsin-Madison.

Kitamura, Y., G. Tripathi, and H. Ahn (2000): "Estimating Parameters under Conditional Moment Restrictions by Smoothing the Empirical Likelihood," Manuscript, Department of Economics, University of Wisconsin-Madison.

LeBlanc, M., and J. Crowley (1995): "Semiparametric Regression Functionals," *Journal of the American Statistical Association*, 90(429), 95–105.

Liero, H. (1982): "On the maximal deviation of the kernel regression function estimate," *Mathematische Operationsforschung und Statistik. Series Statistics*, 13, 171–181.

Newey, W. K. (1985): "Maximum Likelihood Specification Testing and Conditional Moment Tests," *Econometrica*, 53, 1047–1070.

——— (1994): "Kernel Estimation of Partial Means and a General Variance Estimator," *Econometric Theory*, 10, 233–253.

Owen, A. (1984): "The Estimation of Smooth Curves," Stanford Linear Accelerator Center Publication 3394.

——— (1988): "Empirical Likelihood Ratio Confidence Intervals for a Single Functional," *Biometrika*, 75(2), 237–249.

——— (1990): "Empirical Likelihood Ratio Confidence Regions," *The Annals of Statistics*, 18(1), 90–120.

——— (1991): "Empirical Likelihood for Linear Models," *The Annals of Statistics*, 19(4), 1725–1747.

Priestley, M. (1981): *Spectral Analysis of Time Series*. New York: Academic Press.

Qin, J., and J. Lawless (1994): "Empirical Likelihood and General Estimating Equations," *Annals of Statistics*, 22, 300–325.

——— (1995): "Estimating Equations, Empirical Likelihood and Constraints on Parameters," *The Canadian Journal of Statistics*, 23(2), 145–159.

Staniswalis, J. G. (1987): "A Weighted Likelihood Motivation for Kernel Estimators of a Regression Function with Biomedical Applications," Technical Report, Virginia Commonwealth University.

Wald, A. (1943): "Tests of statistical hypotheses concerning several parameters when the number of observations is large," *Transactions of the American Mathematical Society*, 54, 426–482.

Whang, Y., and D. Andrews (1993): "Tests of Specification for Parametric and Semiparametric Models," *Journal of Econometrics*, 57, 277–318.

Wooldridge, J. (1992): "A Test for Functional Form against Nonparametric Alternatives," *Econometric Theory*, 8, 452–475.

Yatchew, A. (1992): "Nonparametric Regression Tests Based on Least Squares," *Econometric Theory*, 8, 435–451.

Zheng, J. (1996): "A Consistent Test of Functional Form via Nonparametric Estimation Techniques," *Journal of Econometrics*, 75, 263–289.

Department of Economics, University of Wisconsin, Madison, WI-53706.
*E-mail address*: gtripath@ssc.wisc.edu

Department of Economics, University of Wisconsin, Madison, WI-53706.
*E-mail address*: ykitamur@ssc.wisc.edu